心理学导读系列

How to Think Straight about Psychology (7th Edition)

与“众”不同的心理学

——如何正视心理学

（第七版）

【加拿大】Keith E. Stanovich 著
范 照 邹智敏 等译
杨中芳 审校



中国轻工业出版社

图书在版编目 (CIP) 数据

与“众”不同的心理学：如何正视心理学（第七版）/（加拿大）斯塔诺威克（Stanovich, K. E.）著；范照等译.
—北京：中国轻工业出版社，2005.2
（心理学导读系列）
ISBN 7-5019-4697-3

Ⅰ.与… Ⅱ.①斯…②范… Ⅲ.心理学 Ⅳ.B84

中国版本图书馆 CIP 数据核字（2004）第 135195 号

总 策 划：石 铁
策划编辑：李 峰
责任编辑：张乃東 孙 琦 责任终审：杜文勇
版式设计：史春雨 责任监印：刘智颖

出版发行：中国轻工业出版社（北京东长安街 6 号，邮编：100740）
印 刷：北京天竺颖华印刷厂
经 销：各地新华书店
版 次：2005 年 2 月第 1 版 2005 年 2 月第 1 次印刷
开 本：850 × 1168 1/32 印张：11.50
字 数：200 千字
书 号：ISBN 7-5019-4697-3/B・040 定价：20.00 元
著作权合同登记 图字：01-2003-8610
咨询电话：010-65595090，65262933
发行电话：010-65141375，85119845
网 址：http://www.chlip.com.cn
E-mail：club@chlip.com.cn
如发现图书残缺请直接与我社读者服务部（邮购）联系调换



中 译 本 序

立意要翻译本书已经是十多年前的事了，当时本书才出第一版，暑假我正在大陆教授研究生“研究方法”的课，发现了这本好书，立刻鼓励一位研究生把它翻译出来，可惜后来因为种种原因，不了了之。像这世界上的许多事情一样，热情过后，这个意念很快就被许多其他更急迫、更切身的事情所淹没。

直到2002年，我在中山大学再次教“研究方法”这门课，用了本书的第六版作为教科书，得到学生热烈的反响，把它翻译出来的念头才再浮现。底稿是由那一年上这门课的4名博士生及8名硕士生草拟。在范照及邹智敏的协调下，经过若干遍的互读、再读、重审（正值原书换版），才到了我的桌上。我在几经艰难的排期下，在广州、珠海、旧金山、香港、北京、台北、青岛；在饭堂里、书桌前、沙发上、病榻上，断断续续地逐字完成了整本书的审校工作。真是非常辛苦，我老了！

虽然说想翻译本书已经有十多年了，但是现在回想起来，当时如果真的翻译成书，其影响力必定不如现在，因为当时人们对心理学的认识或是理解，以及大众媒体对心理学的重视或是炒作，都远远不及现在，书中所涉及的那些利用心理学专家的虚名在电视电台大放厥词，或是利用伪心理学的研究成

果进行招摇撞骗的事例，当时在大陆尚不普遍。人们尚不完全明了有心理学这门学科，自然就不会对书中所述产生共鸣。

现在，则不同了。

通过大众媒体的大力宣传，心理学已经成为一个时尚。任何事件请一个心理学家来点评一下已经变成能够满足人们好奇心及偷窥欲的卖点，从而形成有如八卦杂志一样的“固定销路”。“媒体心理学家”也应运而生，对任何事件或人都可以侃侃而谈地讲出一套心理学的“道理”。对任何个人的任何心理问题，都有一套教其“自助”的方法。通过营销专家的包装，制造出不少畅销的“心理学常识”，建构了一个心理学的文化产业，抢走了不少算命的、看手相的、看星座的人的生意。这些所谓“心理学文化产业”的消费者（读者、观众、听众等）之所以会更听信心理学专家的话，一是因为他们认为心理学本来就与算命的、看手相的、看星座的没有太大差别，只是没有那么俗，上得了学术的殿堂而已。二是因为他们“以为”这些所谓的“心理学专家”所给予的“心理常识”是有科学研究为依据的。

本书清楚地告诉我们。这两种想法都是对心理学的误解。我们必须正视心理学。

心理学是一门科学。心理学的知识是靠科学研究来累积的。现在在大众媒体上，由大部分“心理学专家”所散播的“心理学常识”，都是专家们自己 “个人意见”的表达、“个别经验”的总结或“临床个案”的举例。按本书所定下的“科学研究标准”，都是不合格的心理学知识，因此其信息质量并不高，可信度也并不会比算命的、看手相的、看星座的好到

哪里去。如果发现这些专家讲的“道理”对你很有用，那很有可能是“安慰剂效应”而已（本书详细介绍这一效应）。

为此，本书可以说是一本引导消费者想要用心理学知识的读者如何评价心理学文化产品的购买指南，让想要应用这些“心理学常识”的消费者知道，要如何来判断哪些“道理”是有科学依据的，因此是可信的。而哪一些只是不值一文的“废话”，只有娱乐价值。

从这个角度来看，这本书是适合一般有兴趣的普通大众看的，带有纠正人们心目中对“心理学”之误解的意味，故将书名定为“与‘众’不同的心理学”。我在此也要感谢我的同事程乐华，想出这样一个切题的书名。

从另外一个角度来看，这本书作为心理学系二年级本科生的第一本“研究方法”课的教课书，也是再适当不过的了。因为它贴近生活，学生可以在其日常生活中找到实例，来讨论如何可以做一些真正“科学”的心理学研究，从而更理性地来探索人类心理的奥秘。也因为作者文笔简洁但锋利，诙谐但冷静，逻辑性很强，说服力极高，可以给学生带来很大的脑力振荡及急转弯，增加他们的创新思考能力。

总之，本书对揭去人们对心理学之误解的面纱做了很大的贡献，让读者看清心理学的真面目。我强力推荐这本书给所有想知道心理学是怎么样一回事以及心理学知识是如何累积起来的读者。它指出心理学虽是一门新兴的科学，但是却拥有一个非常古老的传统，因此人们对它的误解也很深，以为对心理学研究所累积的知识，可以不用现代“科学”的标准来衡量及评价。本书则告诉我们，判定心理学知识的好坏，是

和其他的科学一样，一定要用“科学”的标准。书中用了许多精彩的例子说明这些标准是什么，以及要如何用之来分辨真正的科学心理学知识与迷惑大众的伪科学心理学常识。

杨中芳

写于中山大学珠海校区

2004年11月2日凌晨

鸣　谢

本书除了初稿的译者已在各章后署名外，还有许多的中山大学心理学系的研究生及本科生同学们都参与了互校、阅读、找错等的工作，让我们这本书能够译得更准确、更顺畅。因此，值得把他们的名字记下，一并表示感谢。

2001级本科生：丁宇、马春蓉、史松衢、孙静、何妙玲、李洁、李天放、肖凌、肖庐奇、余梦、张浩宇、邹恒、郑茵、胡姗姗、赵冀、黄德斌、黄乐全、谢志成、曾广斌、曾勇军、赖云鹏。

2003级研究生：赵瑾东、何淑瑜、邓咏喻、刘雪琴、陈淑明、董蓓、袁俏芸、章睿健、王亚、崔吉芳、崔丽弦、刘淑华、阮福金。

2004级博士生：杨斌让。

总之，本书是中山大学复系（2001）以来，师生努力的一个成果，我们由衷地感谢同学们的支持与参与。

目　录

序　言

在人类知识的沧海中，有这样一套知识体系，它尚不为大多数人所知，但却关系到人类各式各样的行为和意识，我们可以用它来解释、预测和控制人的行为。懂得这套知识体系的人可以用之来分析其他人，并能够用之更加完整和精确地来理解及决定他们行为和想法的真正原因。

可能会令你惊讶的是这套不为人知的知识体系竟会是心理学！

当我说心理学仍不为人知时，你一定在想，我是在开玩笑吧。因为现在书店里充斥着大量标以心理学为主题的书籍，电视和广播媒体每天频频播放关于心理学的脱口秀，报纸和杂志也开辟有心理学专栏，怎么能说没有人知道心理学呢？然而，就某种至关重要的意义上来讲，心理学确实仍是一个不为人知的知识领域。

尽管心理学似乎得到了很多媒体的关注，但是心理学这个知识体系的绝大部分内容仍不为公众所知。那些通过大众媒体传播的“心理学”知识，在很大程度上只是一种幻象。很多人不知道他们在许多书店里看到的大部分所谓心理学书籍，都是由一些在心理学界根本站不住脚的人所写的；很多人也不知道多数在电视上被贴上“心理学家”标签的那些人，根本

得不到美国心理学会（APA）和美国心理协会（APS）的承认；他们更不知道，大多数出来混个“脸熟”的心理学“专家”，对心理学领域知识的积累未曾做过任何贡献。

传媒对“心理学”话题的短期炒作，不仅向公众传递了许多不准确的信息，而且还模糊了人们对心理学领域中真正的、正在发展的知识体系的认识。公众不能肯定到底哪些是心理学，哪些又不是心理学，也不知道该怎样独立地去评估媒体上流传的那些关于人类行为的言论。更麻烦的是：很多人正利用公众没有评估能力和技术的这一弱点，或是他们认为心理学言论是无从评估的错误观念，在谋取私利。这种认为心理学言论无法评估的谬论，有时也被理解为“心理学没有行规，因此可以百无禁忌地做什么都行”，这种心态对社会危害极大正是本书要讨论的误区之一。很多伪科学正是利用公众的这些无知，大行其道，从中获取暴利，摇身演变成各类价值上百万美元的产业。人们并不知道许多伪科学（例如占星术、通灵手术、超速阅读、生物节律、潜意识自助录音带和通灵侦探等）所宣称的“真理”，其实已经被验证是错误的了。本书所提及的这些伪科学行业，助长了媒体对科学做出耸人听闻报道的趋势。这种趋势对心理学的危害远要比对其他学科的危害更甚，理解其中的原因，是学习如何正视心理学的一个重要环节。

本书不是写给即将成为心理学的研究者们看的，而是为一个更大的读者群——应用心理学信息的人们而写的。本书适用于初学心理学的学生，也适用于通过大众媒体对心理学知识有一些了解、又想知道如何去评价这些信息之合理性的广大读者们。

本书不是一本正规的介绍心理学的入门书，它没有总结心理学领域已取得的一系列研究成果。实际上，要纠正公众那些已经被传媒误导了的对心理学的看法，单靠到大学里选修一门心理学导论的课程可能是不够的，也不是解决问题的最好办法。因为很多对心理学抱有很大兴趣的外行人，他们没有时间、没有钱或是没有机会可以到大学去参加这类正规的学习。更重要的原因是，我作为一名大学心理学课程的教师，不得不承认我和我的同事们在大学里教心理学时，也常常没能很好地引导初学者对心理学这门科学有一个正确的认识。因为初级水平的课程设计中通常没有包括对批判性分析思维技巧的训练，而这正是本书讨论的焦点所在。作为教师，我们常常只会忙于将“研究发现”塞入教学内容当中。每次我们在和学生讨论到诸如媒体眼中的心理学等稍微偏离教学大纲的话题时，都会感到有些内疚，并开始担心自己会不会因为跑题，不能在学期结束前教完自己真正想要讲授的内容。

让我们来看看目前现有的一般心理学导论教科书，它们很多都已经有 600 ~ 800 页那么厚了，其中参考了几百个已出版著作中的研究。当然，含有如此丰富的材料并没什么错，至少它反映了心理学知识在不断增长。但是，这种情况也会带来一些负面效果，教师们常常只忙于塞给学生一大堆的理论、事实和实验，而没时间去回答学生在学习过程中所提出的问题，也来不及纠正他们的错误观念。这主要因为教师们（包括入门类教科书的作者）想当然地认为只要学生接触了足够多的心理学研究，自然就可以从中推断出他们所提问题的答案。他们希望学生可以从对心理学各领域研究的陈述及讨论中，自

行挖掘出隐藏其中的各类问题的答案。但是这个希望往往都落空了，到这门课的最后复习阶段，或是在学期结束前，教师们才无比震惊和沮丧地发现，学生提出的一些问题及说法，应该是他们在课程开始第一天就应该提出来讨论的，而不是最后一天。比如有学生提出：“既然心理学实验不同于实际生活，那么这些实验能告诉我们什么呢？”；“心理学可以像化学那样成为一门真正的科学吗？”；“可是，我在电视上听一位临床心理学家所讲的，好像正好与我们教科书上说的相反”；“我认为这个理论毫无价值——因为我弟弟的行为和这个理论所说的截然相反”；“心理学不过是些一般常识而已”；“每个人都知道什么是焦虑——为什么还要费工夫去定义它呢？”；“心理学只不过是一些不同看法的汇总而已，不对吗？”对于很多学生而言，这些问题是没办法仅靠思考书中的内容来找到答案的。在本书中，我将明确地回答上述这些问题，并澄清这些说法背后的种种观念上的混乱。

其实，许多研究都已经发现，心理学入门课程通常不能很好地纠正初学者对心理学存有的众多误解（Best，1982；Higbee & Clay，1998；McCutcheon，Furnham，& Davis，1993；Vaughan，1977）。有一位叫沃恩（Vaughan）的研究者说：“我不得不说，一般那些‘导论’性质的课程，事实上，对纠正学生已有的错误观念作用不大”（Vaughan，1977，p.140），而且，他还进一步得出结论：“目前我们在学生中间还没有培养出一种批判性思考的氛围，这种氛围可以让学生对书中没有讲清楚的地方提出疑问”（p.140）。沃恩后面的这结论与本书撰写的初衷相吻合，心理学也许比其

他科学更需要批判性的思维来引导学生将科学以外的杂草加以铲除。这种批判性思维是学生能够独立评价心理学信息所必须掌握的！

即使若干年后学生们不再记得心理学入门课程中的内容，但他们仍然可以运用本书所涵盖的基本原理去评判有关心理学的其他言论。就像学生们即使把埃里克森（Erikson）的人生发展阶段论完全忘记了，他们也仍可以运用本书介绍的思维工具，去辨别在媒体中不断涌现的心理学信息的真伪。一旦掌握这些可以终身受用的思维技巧，就有助于我们去评判各种理论主张。首先，它将使我们能够对那些看似合理的事实做出一个初步、总体的判断。其次，这些技巧可以作为评估“专家”观点可信度的标准。在这个复杂的社会里，人们越来越依赖专家的观点；在获取知识时，判断专家观点的可信度就变得尤为重要了。虽然这些批判性思维技巧可运用于各个学科或整个知识体系，但在心理学领域里它们显得尤其重要，因为大众媒体常常歪曲这个领域的内容。

对任何试图阻止心理学被曲解的努力，许多心理学家都抱着悲观的态度。虽然这不是完全没有理由的，但是，如果我们就此放弃，这种曲解就会越来越严重。我这本类似“消费者指南”式的书，就是想要表明，我们不能让这样一个恶性循环，成为缠住心理学的魔咒。

尽管我很高兴这本《与“众”不同的心理学》能够陆续出版了这么多次，但有些遗憾的是，本人当初撰写本书第一版的理由及目的，至今都依然没有改变。媒体对心理学的介绍仍然像以前一样在误导大众，而学生在开始上心理学入门

课程时，还是带着与以往同样多的，对心理学的误解。这导致本书以后几个版本之出版理由与目的，都和第一版一模一样。不过令人欣慰的是，我的写作初衷现在已经得到越来越多心理学教师的回响。斯坦福大学心理学家罗杰·谢泼德（Roger Shepard）表达了与本书写作初衷相同的看法：“虽然大多数心理学的本科生也许不会走上从事科学研究的道路，但我们仍然希望他们有能力去评估那些片面、幼稚、混乱及夸张的报导，这些报导不断地出现在媒体上，自称是社会科学的‘新发现’……那些认为可以通过未经验证的常识，或者更糟糕的，可以通过星相学之类无实验根据的伪科学，就能达到对人类行为和心理现象的充分了解之谬论，势必将会不断地向我们提出挑战”（Shepard，1983，p.855）。

本书的目的是对批判性思维技巧作一个简略的介绍，这些技巧将帮助人们更好地去理解心理学的论题以及他们周围世界所发生的事情。

第七版更新内容

第七版的《与“众”不同的心理学》在结构上没有作很大的改动，因为第五版已经对其中一个章节作了较大的重组。各章节的内容和顺序也保持原样。应评审者和读者的要求，这一版作了一点“瘦身”。读者和使用者都不希望本书篇幅增大，因此，实际上，这一版没有像以前的版本那样增加篇幅，第七版反而比第六版缩减了15%左右。篇幅虽然缩减，但没有删减任何概念，只是删掉了一些冗余拖沓的例子，保留下最

好的（依据评审者和读者的反馈意见来做的）。

最重要的是，我更新和重写了书中的许多例子，用最近的研究和论题替换掉了那些过时的事例。我也极力引用与本书所提及的各种概念和实验相关的最新研究资料。因此，在这一版中总共有129处新引文，读者在此可以得到所有例子和概念的最新参考资料。

本书出版的宗旨一直保持不变——对批判性思维技巧作简略介绍，帮助学生更好地去理解心理学的论题。在20世纪90年代，大学里强调批判性思维技巧的呼声越来越高（Halpern，1998）。的确，有些州立大学系统已经进行了加强批判性思维教育的课程改革。与此同时，也有其他教育学学者认为，批判性思维技巧不应该脱离特定的学科内容。而《与“众”不同的心理学》正好融合了这两个观点，帮助教师在教授丰富的现代心理学知识的同时，传授批判性思考的技巧。

欢迎读者把你们对本书的批评意见寄到下面的地址和我交流：Keith E. Stanovich，Department of Human Development and Applied Psychology，University of Toronto，252 Bloor St. W.，Toronto，Ontario，Canada，M5S 1V6. E-mail: KStanovich@oise.utoronto.ca

（郭妙燕　译）

第一章　充满活力的心理学

——在科学阵营里还干得蛮好

弗洛伊德盛名之累

如果我们在大街上随机拦截100个行人，请他们说出任何一个已逝或在世的心理学家，那么像韦恩·戴尔（Wayne Dyer），和其他一些曝光率很高的"媒体心理学家"，是一定会被提及的。假若我们把这类心理学家排除在外，让行人以对心理学做出公认贡献为标准，列出心理学家来，那排在首位的一定是西蒙·弗洛伊德（Sigmund Freud），而B. E. 斯金纳（B. E. Skinner）将可能以较大差距屈居第二。至于其他的心理学家，人们很可能就再怎么想，也想不出了。由此可见，是弗洛伊德，加上在媒体上经常出现的通俗心理学家，共同塑造了公众对心理学的印象。

"声名遐迩"的弗洛伊德，不但深远地主宰了公众对心理学的认识，而且也导致公众对心理学产生了许多误解。例如，很多上"心理学导论"课的学生都惊讶地发现，在美国心理协会和美国心理学会这两个正式的学术组织成员中，把关注弗洛伊德精神分析学派的心理学家加在一起，分别也只占了10%和5%。

事实上，现代心理学研究完全没有被弗洛伊德的思想所左右，更没有被他的想法所限定。只有媒体和一些人文学科领域才会这么认为。在种种现代心理学研究的论题、数据和理论中，他的工作只不过是沧海一粟罢了。心理学研究和理论包括了5位近期诺贝尔奖获得者的工作：他们是大卫·休伯尔（David Hubel），丹尼尔·卡尼曼（Daniel Kahneman），赫伯特·西蒙（Herbert Simon），罗杰·斯佩里（Roger Sperry）和特斯滕·维瑟（Torsten Wiesel），另外包括美国国家科学基金会的前负责人里查德·阿特金森（Richard Atkinson），而这些人都不为公众所知。

更不幸的是，心理学与弗洛伊德的过分认同，给这门学科带来的是夸张和曲解。其中最为糟糕的是，弗洛伊德的研究方法已经远远不能代表现代心理学研究所运用的研究手段了（别忘了弗洛伊德开始他那些知名的研究工作是一百多年前的事了）。因此，对弗洛伊德研究方法的理解，会造成对心理学研究的极大误解。首先，弗洛伊德不做实验控制，而我们从第六章的内容可以清晰地了解到实验控制是心理学研究最有力的武器；第二，弗洛伊德通过个案研究来证实或证伪理论，而我们可以在第四章中看到这种方法的谬误之处；最后，弗洛伊德的研究最大的问题，在于他的理论与行为数据之间的关联。我们同样可以在第二章的内容中了解到一个理论要达到科学标准，该理论与行为数据间的关联至少要满足几个条件。而弗洛伊德的理论却完全不能满足这些评判准则（Crews，1996，1998；MacMillan，1997；McCullough，2001；Watters & Ofshe，1999）。简言之，弗洛伊德是根据个案和内省数据资

料建立起他的精致理论的，但是，这个理论的数据基础并不牢固。他只是集中精神于建立复杂的理论构架，但是却没有像现代心理学家那样，确保这些理论构架是建立在可靠、可重验的行为数据上。因此我们可以说，精通弗洛伊德的那套研究模式，对理解当代心理学的研究方法而言，反而是个阻碍。

本章中将从两个方面来处理这个所谓由弗洛伊德所带来的盛名之累：①我们通过阐释现代心理学的多样性，能够很清楚地看到弗洛伊德的理论在这一学科中，事实上是处于相对次要的地位（见 Haggbloom et al.，2002；Robins，Gosling，& Craik，1999，2000）；②我们将探讨在心理学领域中，大部分的研究所共享的研究方法有哪些特点，它们与过时的、弗洛伊德式的知识累积方式，是完全不同的。以此来澄清公众对现代心理学研究方法的误解。从而让公众看到他们对弗洛伊德研究的认同，让他们忽略了能整合现代心理学的惟一特性，那就是：用科学的方法来理解人们的行为。

现代心理学的多样性

现代心理学包含了多种多样的内容和层面，这种多样性多少也使得整个心理学显得松散或不够统一。正如美国心理基金会杰出教学奖得主亨利 · 格雷特曼（Henry Gleitman）对心理学所做出的描绘："一个松散地联系在一起的学术王国，它横跨了从生命科学为一端到社会科学为另一端的所有领域"（Gleitman，1981，p.774）。

理解了心理学令人惊叹的广泛性以及调查手段的多样性，

对于真正地理解心理学这一学科的本质是至关重要的。从下列的事实中可以清晰地看到这一点。美国心理学会（APA）包括53个分支，每个分支代表一个特定的研究或应用领域（见表1.1）。从表中我们可以看到心理学家研究的主题范围、研究背景跨度和所使用的不同的研究角度。同样，另一个美国心理协会（APS）也是分支众多。一般来讲，从表1.1中可以了解到心理学每个分支都是一个专门化的领域。事实上53个分支中的每一个，又都包括了相当广泛的研究领域和子分支，以至于很难找到不被心理学所包含的领域。

表1.1　美国心理学会（APA）的分支机构

1. 普通心理学（General Psychology）
2. 心理学教学（Teaching of Psychology）
3. 实验心理学（Experimental Psychology）
5. 评价、测量和统计（Evaluation，Measurement，and Statistics）
6. 神经行为科学和比较心理学（Behavioral Neuroscience and Comparative Psychology）
7. 发展心理学（Developmental Psychology）
8. 人格和社会心理学（Personality and Social Psychology）
9. 社会问题的心理学研究（Psychological Study of Social Issues）
10. 心理学和艺术（Psychology and the Arts）
12. 临床心理学（Clinical Psychology）
13. 应用咨询心理学（Consulting Psychology）
14. 工业和组织心理学（Industrial and Organizational Psychology）
15. 教育心理学（Educational Psychology）
16. 学校心理学（School Psychology）
17. 理论咨询心理学（Counseling Psychology）
18. 公共服务中的心理学家（Psychologists in Public Service）
19. 军事心理学（Military Psychology）
20. 成人发展与老龄化（Adult Development and Aging）
21. 应用实验和工程心理学（Applied Experimental and Engineering Psychology）
22. 康复心理学（Rehabilitation Psychology）

23. 消费者心理学（Consumer Psychology）
24. 理论和哲学心理学（Theoretical and Philosophical Psychology）
25. 行为实验分析（Experimental Analysis of Behavior）
26. 心理学史（History of Psychology）
27. 社区心理学（Community Psychology）
28. 精神药理学和药物依赖（Psychopharmacology and Substance Abuse）
29. 心理治疗（Psychotherapy）
30. 心理催眠（Psychological Hypnosis）
31. 各州心理学常务联会（State Psychological Association Affairs）
32. 人本心理学（Humanistic Psychology）
33. 智力缺陷和发展性障碍（Mental Retardation and Developmental Disabilities）
34. 人口与环境心理学（Population and Environmental Psychology）
35. 女性心理学（Psychology of Women）
36. 宗教心理学（Psychology of Religion）
37. 儿童、青少年和家庭服务（Child, Youth, and Family Services）
38. 健康心理学（Health Psychology）
39. 心理分析（Psychoanalysis）
40. 临床神经心理学（Clinical Neuropsychology）
41. 心理学和法律（Psychology and Law）
42. 独立从业的心理学家（Psychologists in Independent Practice）
43. 家庭心理学（Family Psychology）
44. 男女同性恋的心理学研究（Psychological Study of Lesbian, Gay, and Bisexual Issues）
45. 少数民族的心理学研究（Psychological Study of Ethnic Minority Issues）
46. 媒体心理学（Media Psychology）
47. 锻炼和运动心理学（Exercise and Sport Psychology）
48. 和平心理学（Peace Psychology）
49. 团体心理学和团体治疗（Group Psychology and Group Psychotherapy）
50. 成瘾（Addictions）
51. 男性和男性化的心理学研究（Psychological Study of Men and Masculinity）
52. 国际心理学（International Psychology）
53. 临床儿童心理学（Clinical Child Psychology）
54. 幼儿心理学（Pediatric Psychology）
55. 药物疗法（Pharmacotherapy）

注：没有第 4 和 11 分支

多样性的意义

学生们在来学习心理学时，都希望学到一套统一又完整的心理学理论，可以用它来解释人类意识和行为的所有方面。有这种想法的学生往往会大失所望。因为他们所能学到的不是一个完整统一的大理论体系，而是许多不同的小理论，每一个只涵盖行为的一个有限层面（Griggs，Proctor，& Bujak-Johnson，2002；Zechmeister & Zechmeister，2000）。对于跨度如此之大的心理学，建立一个大统一的理论基础确实是相当困难的。实际上，许多心理学家都认为这种建立大统一理论体系的目标是不可能达到的。尽管如此，还是有一些心理学家在某些领域里，试图从事整合的工作（Gibson，1994；Kenrick，2001；Kimble，1994；Solso & Massaro，1996）。例如，在过去10年中，由于进化论心理学家的理论贡献，心理学的内部凝聚性大大增强。这些学者试图通过将人类的心理过程视作为一系列服务于宗族识别、配偶选择、合作、社会交换、儿童抚养等（Barrett，Dunbar，& Lycett，2002；Bjorklund & Pellegrini，2000，2002；Buss，1999，2000，2001；Cosmides & Tooby，1992，2000；Pinker，1997，2002），对进化功能而言至关重要的关键机制，从而实现了这些心理过程在概念上统一起来的可能性。

但无论心理学家各自站在什么立场，他们基本上都同意建立大统一理论框架是非常困难的，即使这真的有可能，那也得很多很多年以后才能实现。缺乏理论的整合，不仅使学习心理学的人失望，同时也为许多否认心理学能为科学做出贡献

的批评者提供了把柄。他们的批评往往源自一个错误观念：所有真正的科学都应该具有一个宏大的、统一的理论根基。这个观念的错误之处在于，它忽视了很多其他所谓科学也同样缺乏一个大统一的概念体系。正如哈佛大学心理学家威廉·艾斯蒂斯（William Estes）所明确指出的：

实验心理学家们所面临的困境，在整个科学领域并不鲜见。物理学在20世纪早期就是处于分裂状态，即使本科教学也是同时讲授不同的观点，以至于我是从机械学、热力学、光学、声学和电子学等不同的课程中去掌握物理学的。同样化学也曾被分为无机化学、有机化学、物理化学和生物化学。当时这些学科之间的交叉融合并不比现在的心理学分支少。物理和化学都是在出现了更高抽象水平的数学理论后才实现整合的。与心理学类似，医学现在也被分为很多分支，也还没有出现大统一的迹象（Estes，1979，pp.661～662）。

划分一个学科的领域范畴往往是历史的偶然，与学科本身没有太大关系。例如，大学的院系很多时候并不是按照知识本身的结构来设置的，而是按行政管理上的方便来划分的。因此院系划分也是可以改变的。正如前美国心理学会（APA）主席威廉·彼万（William Bevan）所讲，心理学理论本身四分五裂，而心理学系居然还能存在下去，恐怕更多是出于管理上的方便，或者只是为了能让系内成员方便地接收自己的邮件罢了。

可以设想一下，假如我们废除心理学系，而把它的成员分到其他相关的系里（Gardner，1985；Scott，1991；

Spence，1987）。生理心理学家可以分到生物系（Wilson，1998）；社会心理学家可以分到社会学系；认知和知觉心理学家可以分到认知科学系（Gardner，1985）；组织和工业心理学家可以分到商学院；临床和咨询心理学家可以分到社会工作系、人力资源系和教育系；发展心理学家可以分到教育系、认知科学系或人力资源系等等。分到各系后，心理学家几乎感觉不到和新同事之间在学术交流方面会与旧同事有什么大的差异，而且往往会觉得与新同事更容易合作共事。由此可见，在内容上现代心理学的各领域根本很难整合在一起，成为一个彼此相互关联的整体。所以也就需要能够在更高的水平上去寻找可以整合心理学这一学科的支撑点。

一旦我们理解了决定学科结构的社会和历史因素，我们就可以认识到，要求所有领域完全统一整合并不合乎逻辑。所以“心理学研究”这个术语在某种程度上比“心理学”更能准确反应这一学科的多样性。这个用法还能让学生更容易理解，为何在解释和预测行为方面，一些心理学领域会比另一些领域取得更多及更重大的科学进展。而“心理学”这个词就表达不出这样的差异。反而让人感觉它是一个具有整合性的学科，这一整合性当然正是心理学这一学科所没有的。如果我们试图去寻找心理学这一学科的整合性，那一定不要在心理学家所研究的课题内容的相互关联性上去找。相反地，我们应把重点放到心理学家发展形成新知识的方法上去找。在这方面，是我们惟一有可能在心理学家中间找到共同点的地方。然而也正是在这方面，是我们发现心理学被误解最多的地方。

在科学上达到一统

用科学的方法了解人类的行为，是心理学各分支之间仅含的连贯性。事实上，心理学如果说有任何特殊性，那就在于它是在广泛地使用各种科学方法，来研究人类的行为。仅说心理学是研究人类行为的，并不能把心理学与其他学科分开来。其他专业机构和学科，像经济学家、小说家、法律、社会学、历史、政治学、人类学和文学研究等，也都在某种程度上与人的行为相关。在这一点上，心理学并非与众不同。

在实际应用方面，心理学研究也并非与众不同。例如，想帮助别人的出发点促使很多大学生主修心理学。但帮助别人也是很多其他学科的一部分，如社会工作、教育、护理、职业咨询、理疗、司法学、人力资源和语言治疗法（甚至哲学——见 Marinoff，1999）等。同样地，通过咨询帮助他人也是教育、社会工作、警事工作、护理、神职工作、职业治疗等学科的手段。因而，如果只是培养专业人员，通过咨询的方式来帮助别人，并不一定需要开辟一门叫“心理学”的学科。

其实只有两件事可以让心理学有成为是一门独立学科的理由。那就是，用科学的技术手段来研究人或非人的所有行为，以及把这些研究成果在付诸应用时，也通过科学的方法来达成。如果不是这样，那么心理学也就失去了存在的意义。

不过，有别于其他行为学科，心理学试图给公众两方面的保证。其一是心理学研究所得的有关行为的结论是从科学实征中得来；其二是心理学的实际应用都源于科学方法。心理学在

发展过程中是否曾经偏离了这两个目标呢？回答是肯定的，而且经常如此（Dawes，1994；Fox，1996；Garry，Frame & Loftus，1999；Lilienfeld，1998，1999；Loftus & Guyer，2002；Mook，2001；Watters & Ofshe，1999）。本书就是想帮助我们掌握如何可以更好地达到这两个目标的途径。在第十二章里，我将会再回来讨论心理学家如何因罔顾科学的标准，而将自己的合法性打了折扣。但是，原则上讲，正是这些标准使得心理学有资格成为一个单独的领域。如果有一天心理学真的认为这些目标不值得追求——因而不想再坚守科学的标准——心理学则大可以关门大吉，让各分支按内容归并入其他学科，因为那时它会变成是一个完全多余的学科。

从这一点可以很清楚地看到，任何人要理解心理学，第一要知道的就是，心理学之所以为心理学，就是因为它是一门以数据为根基的行为科学。让人们理解到这一点，也是贯穿本书的主旨。因为惟有认识了这一点，我们方才有能力去想清楚心理学是什么。相反地，人们之所以对心理学产生这么多误解及困惑，也是因为他们不认可心理学为一门科学。我们不是经常听到圈外人的这类质疑吗？为什么会这样呢？

让公众觉得心理学不是一门科学有多方面的原因。我们会在随后的章节中讨论到，人们对心理学到底是什么的困惑，多半是由那些伪心理学教师一手精心炮制的。我们的社会中已经发展出相当一批伪科学信念系统的发明者。他们因为自身的利益，总是迷惑大众，把心理学说成是可以"信口雌黄"得到的，并且宣称并不存在任何理性的标准来评价心理学的言论及主张。于是，在这种氛围影响下，市面上鼓吹类似"催

眠瘦身法”，“激发你的潜在心理能量”和“梦中学法语”等的东西，以及其他营业额高达数百万元的“心理自助”* 产业就大行其道了。实际上，这些东西不是毫无科学根据，就是与现有证据相抵触。

另一个对科学心理学产生抗拒的力量，是来自一股不想让科学进入原本是由“想当然的权威”和“一般常识”所支配的领域。历史上提供了许多事实，显示公众对科学知识是有抗拒的。也就是说，公众更倾向于用哲学的思辨、神学的预言或是民间的常识来解释世界。每个学科在其发展的过程中，都曾经过这一抗拒阶段。例如，与伽利略同一时代的其他学者，居然拒绝使用他发明的新式望远镜来观测星空，因为有环绕着木星运行的卫星存在，这一事实冲击了他们的哲学和神学信念。几个世纪以来，人体解剖学的知识发展缓慢，就是因为它受到公众及教会对解剖尸体之禁忌的影响（基督教认为身体的内部是神的领地，见 Grice，2001）。查理 · 达尔文（Charles Darwin）之不断受到重斥；保罗 · 布洛卡（Paul Broca）的人类学会之在法国受到排挤，也都是因为有关人类作为一种生物的知识，在当时被认为对国家具有颠覆作用。

科学的每次重大进步都曾经历过巨大的阻力。不过当人们理解到新的发现不但不会抵毁人性，而且能因拓展我们的知识而更加成全人性时，反对的声音也就消失了。现在谁还会认为星云图及天文理论会摧毁我们对宇宙的赞叹呢？谁会选择尸体解剖开禁前的医疗保健系统，而不选择现在在我们自己社区

* 心理自助是指通过阅读或聆听一些包含临床的案例、治疗步骤、激励性的书籍、录音及录像带，让读者可以在家中自行解决自己的心理问题。

内所获得的良好医疗保健呢？所以说，对星际或人体要做实征研究，并没有诋毁人性。近些年，达尔文的进化说可以说是为遗传学和生物学的惊人进步打下了坚实的基础。然而，即使是在当今我们因为他的学说而逐渐逼进了解人类的本质和起源的时候，残存的、反对它的声音依然存在。在美国，极端教派继续在公立学校鼓吹上帝创世说。研究表明仍有相当大比例的公众（在某些调查中，甚至是超过半数人）不接受人是由“生存竞争、自然淘汰”进化而来的这一事实。如果摆着这么多科学证据的进化生物学都还不断地受到公众的抵制，那我们怎能奢望人们不会对刚发展起来的、目的在对人们长期持有的、有关自己的信息，进行科学检验及挑战的心理学，产生质疑呢？

最后要指出的是，不承认心理学是科学的人往往对什么是科学的认识也是一团糟。在大学里，几乎每一个心理学教师都会遇到一些选择主修心理学的一二年级学生说，他们之所以选择心理学是因为他们不喜欢科学。当教师告知他们心理学正是一门科学时，他们都感到非常吃惊意外。教师对他们接下去的沮丧反应（“我不能相信我还要学统计！”），早已见怪不怪了。当教师再问：“你们在入大学之后选修过生物学和化学吗？”，他们的回答一定是：“哦，当然没有啦！我一直都在回避理科的课程哪！”这里，我们看到的是一个极具讽刺性的现象：学生在对科学一无所知的情况下，居然就确信心理学不是一门科学！

那么，科学到底是什么

要明白心理学是什么，我们必须先明白科学是什么。不过，要明白科学是什么，也许我们还得从科学不是什么谈起，以便澄清大量对这个词的误解。首先，科学不是由研究课题的内容来决定的。对宇宙的任何一方面的研究都可以发展成为一门科学，当然也包括对人类行为的研究。因而，我们不可能把宇宙分为：有的是“科学”的，和有的“不是科学”的课题。虽然历史上确实有过一股强大的势力想把人类排除在科学研究的对象范围之外，但最终都以失败告终。当前针对心理学是否是科学所提出的怀疑，应该还是这场历史争论的余音。有关这场争论，在下面会提到。

另外，科学也不是非得用实验仪器来标榜。科学研究决非是以白大褂、试管、计算机、电子设备、脑科学技术为特征。（即便用这些指标，心理学也完全符合科学的定义，因为几乎在所有大学的心理学系都装满了计算机、化学制剂和各种电子设备。）这些仅仅是科学的附着物，完全不是其本质特征。科学是一套理解宇宙的思考、观察方式及程序，它让我们能更进一步地去了解宇宙的运作。

本章的其余部分，将讨论科学的三个紧密相连的核心特征：①系统的实征主义；②产生大家共有的公开知识；③对可解决的问题之探研。虽然下面将对这三个特征分别进行阐述，但请紧记，这三者是紧密联系地构成一个整体结构。（更详尽的讨论请参阅参考文献中 Bronowski、Cournaud、

Dawkins、Medawar、Popper、Raymo 和 Sagan 的专著。)

系统的实证主义*

用任何一本词典来查找实证主义一词，你都会得到其意为“基于观察的做法”的解释，即科学家是通过探究世界来发现世界的。也许这在你看来，已是不争的事实，这正好说明，科学态度的渊远流传，是几百年的历史了。在此之前，人们对科学的认识并非如此清晰。回想在那个拒绝使用望远镜的年代，人们一直以为知识只有通过冥想或从权威那里才能得到。当伽利略宣布在木星周围观察到了卫星时，另一个学者弗朗西斯科·希兹（Francesco Sizi），不是用观察，而是试图用以下言论来反驳他：

人的头上有七窍：2 个鼻孔、2 只耳朵、1 双眼睛和 1 张嘴；天上的星星也有 2 个吉星、2 个灾星、2 个可问卜的，还有 1 个不知可否的水星。从很多类似的现象，如 7 种金属等，我们可以料定行星的数量也必定是 7……另外，犹太人和其他古老民族，都跟现代欧洲人一样把每个星期分为 7 天，并用来分别命名 7 颗行星；现在如果我们增加行星的数量，则整个系统都将垮掉……还有就是卫星是肉眼见不到的，也就不会影响地球，既然没有用，也就没有存在的必要。（Holton & Roller，1958，p.160）

* 实证主义原文为 empiricism，是指用能观察到的东西为研究对象的主张。大部分我国书籍翻译为实证主义。但因实证主义常也是 positivism 的译词，为避免混淆起见，故选用实证主义一词。

引述上面的这段文字不是为了说明它的愚蠢可笑，而是想指出这便是曾经作为反驳实际观察数据而想出来的、在当时看来相当合理的辩词！我们对它的轻视和嘲笑都因为我们是事后的诸葛亮。经过3个世纪实征方法的洗涤，我们比希兹处于优势。但如果没有这些年洗涤的经验，我们也很可能会频频点头地赞同他。可以说，实征研究方法从来不是那么显而易见的，这就是为什么即使在科学占主导的当今社会，仍有必要教导人们认识它的原因。

不过仅靠纯粹、单一的实征主义是不够的，这一点可以从标题“系统的实征主义”中了解到。观察法是好的而且是必要的，但是单一、非结构的观察不会发现科学知识。例如，你可以从起床到就寝做整日的观察记录，随后你会拥有大量事实，但是这并不会加深你对世界的认识。而科学观察是系统化的，因为它是有结构地去观察，所以其结果能够揭示世界运行的规律。科学的观察通常是为了检验理论；它是想透过测试各种不同的说法，来理解我们生活的世界。科学的这一结构性可以令我们根据观察的结果，去支持一些理论，及否定另一些理论。

可公开验证的知识：重验性与同行评审

从某种特殊意义上讲，科学知识具有公开性，但这公开性不一定就是要把科学发现贴在社区中心的公告牌上，而是指科学知识不只存在于个别人的头脑中。一个基本共识是：任何一个科学发现在没有接受科学界其他人的批评和验证之前，都不能称其为科学知识。知识如果是“特别”的，亦即是一个人想出

来的，而且不受他人的审视及批评，那它一定不是科学知识。

科学是通过重验这一机制来实现其可公开验证性。一项发现如果要在科学界得到公认，它必须要能让其他科学家重复相同的实验，并得到相同的结果。只有这样，我们才说该发现是被重复验证了。所以，这种重复的验证机制保证了某个发现不是因为个别研究者的倾向或错误所造成的。总之，某发现要得到科学界的接受，其前提一定是：可以被最初发现者以外的其他研究者所重验。只有经历了这样的检验，此发现才具有公开性。此时它不再是最初研究者的私有财产，而是可以为其他研究者继续发展、批评及应用到他们自己的领域中去。

诗人约翰·唐（John Donne）讲过"没有人是一个孤岛"，相应地在科学上也没有任何研究者是一个孤岛。每个研究者都与科学社团、已有的知识基础紧密相连，是这种相互连接促进了科学的累积发展。在科学进程中，研究者经常要在先前的知识基础上从事新的探索和超越，这样，就必须保证这之前的知识的累积方法的确是可以作为进一步研究的基础。

可公开验证的知识，是指可以被科学界任何其他人经过重验、批评和继续发展的这个验证程序，而得到的知识。这个标准不仅对科学家是重要的，它也是作为消费者的普通大众应该用来评价出现在媒体中的科学信息的重要依据。正如在第十二章中将要讲到的，骗子或那些施用伪科学的人，与真正科学家的一个重要区别，在于前者往往绕过科学界的常规，不通过科学出版物来发表其研究发现，而是直接通过媒体公布他们的"发现"。在相信那些宣称"科学"的发现之前，一条百颠不破的评判其真伪的方法是先问一下，这些发现是否已

经在具有同行评审制度的正式科学期刊发表了。对这个问题的回答，立刻就可以把伪科学和真理分开来了。

同行评审一般是这样进行的：投稿文章首先提交给若干科学家来评审，评审人将其批评意见提交给一名编辑（通常是该刊物涵盖领域中，在某方面有多年经验的科学家），再由编辑来决定是否发表该文，或需要进一步的实验或统计分析，还是存在某个致命问题和缺陷而被拒绝采用。一般来说，大部分刊物在其内页都会刊登有关该刊物的编审政策及章程，通过它们我们可以获悉此刊物是否执行同行评审制度。

当然，不是所有信息被同行评审过的内容都一定是正确的，但至少它们达到了同行评审的标准。这也只是最低标准而已，而不是最严格的，因为许多学科领域都有各种不同质量的刊物。大部分的科学想法，只要达到最基本的标准，一般都可以在各种各样的刊物上公开发表。因此声称只有范围极窄的数据和理论才能被科学刊物发表的说法其实是不对的。但江湖术士们却常常借此，让媒体和大众相信他们是被“正统科学圈子”的阴谋所害，堵死了他们出版之路。让我们看看像心理学这样的科学领域，提供了多少合法出路吧，《心理学摘要》中罗列着上千种学术期刊的文章摘要，而这些刊物绝大多数是有同行评审的制度的。因此，几乎任何稍微经得起考验的理论和实验都可以找到发表的地方。

我再次声明，我并不是说在《心理学摘要》中发表的发现都是好的或正确的。正如我前面所说的那样，它们仅仅只是满足了最低标准。但是我们倒是可以说，一个假说、理论、说法或疗法如果不能在学科内部经过同行评审发表出来，

就显示它是有问题的。尤其当在证据不足的情况下，又要通过媒体来宣传自己的说法，那么恰恰显示它肯定是伪科学。

虽然在不同学科间同行评审的方式不尽相同，但是其根本逻辑是相同的。同行评审和重验机制是科学保证其客观性和公开性的根基。思想和实验只有经过了其他人的严格审视及尖锐批评，才能开始接受公众验证的洗礼。同行评审机制虽然还未能达到完美，但是它是我们消费者仅有的保障机制。任何忽视或轻视它的做法都会让科学和伪科学同流合污，并受制于它们的淫威（见第十二章）。在随后的章节中，我们将会更详细地讨论，我们不按科学心理学程序实践，忽略评审监督机制，所给我们带来的巨大危害和惨痛教训。

能用实征方式解决的问题：科学家对具可验证性之理论的探求

科学是用来解决可以解决的、具有明确性的问题。也就是说，科学家所解决的问题是那些有可能用现有的实征技术找到答案的问题。如果一个问题是不可能有答案的，或是一个理论是目前没有技术可以来验证的，那科学家就不会去研究它了。例如，“在照顾3岁以下的小孩子时，给予结构化语言刺激的孩子，其日后的阅读学习水平是否会比不给刺激的孩子要好呢？”，像这样的问题就是一个科学问题，因为这个问题可以利用现有的实征方法来获得解答。而“人性本善还是本恶呢？”和“生命的意义是什么？”，则不是可实征的问题，也就不属于科学的问题了。

科学的进展通常是通过先提出理论假说来对世界上某特定

现象进行解释开始，经过由理论推展出预测、再经实验来验证，最后到根据实验结果修正理论，这一连续过程来进行的（理论>预测>实验>修正）。所以科学家所说可解决的问题也就是指“可验证的理论”。理论怎么样才是可被验证的呢？首要的是理论必须对自然界中的特定可观察事件有一定的作用；这就是所谓的可被实征验证性。可验证的标准就是在第二章提到的可证伪标准。

我们知道，对科学只能解决可被实征验证的问题这一点，存在两大误解。第一，科学家经常被刻画成仅仅认为科学问题才是重要问题的形象。当然这并不一定是实际情况，仅仅因为科学家在其专业工作中不解决非科学问题，并不能说明他们对其他非科学的工作漠不关心。

第二，我们说科学家只关注可用实征方式解决的问题，并非意味着所有的问题可以被分类为是可以解决的和不可解决的，或这种划分永远不可以改变。恰恰相反，现在解决不了的一些问题随着理论和实征技术的发展，有一天会成为可解决的。例如 10 年前历史学家都不相信，托马斯 · 杰佛逊（Thomas Jefferson）的奴隶萨莉 · 海明斯（Sally Hemings）抚养的孩子是否是杰佛逊的这一争论，是个可以用实征方式来解决的问题。可是，1998 年这问题已经随着遗传技术的进步得到了解决。刊登在《自然》杂志上的一篇文章（Foster, et al., 1998）说杰佛逊非常有可能是爱斯顿 · 海明斯 · 杰佛逊（Eston Hemings Jefferson）的父亲。

这一例子也说明了一般科学发展和新科学的产生过程。对什么是当前可解决或不可解决问题的争论有很多，科学家自己

通常也因这类问题的模糊性，而难以达成共识。所以虽然科学家们对“只研究可用实征方法来解决的问题”这一标准认识一致，但在具体到底怎么决定哪些问题是可解决的，哪些是不可解决的问题上却争论不休。诺贝尔奖获得者彼特·米德瓦（Peter Medawar）就曾把他的一本书定名为《可以解决的艺术》，书中他指出，科学的部分创造性就体现在，它是在人类知识的最前沿，探索可以用实征技术进行研究的问题。

心理学里有很多很好的例子，说明一个问题是可以从不可解决的向可解决的方向过渡的。很多问题，如“孩子是怎样学会其父母说的语言的？”、“我们为什么会忘记我们曾经记得的事情？”、“加入一个群体后，人们的思想和行为是如何改变的？”。在有人意识到它们可以通过实征方法来解决这些问题之前，人们早就对这些问题提出自己的揣测。随着这种可以用实征方法来解决问题之认识的发展，各个领域中有关人们行为的各项问题，才被集中起来组合起来构成现在的心理学。心理学的问题，也逐渐从哲学中独立出来，发展成为一门实征的学科。

认知心理学家史蒂文·平柯（Steven Pinker）曾讲过，“无知”可以分为问题（problems）和谜团（mysteries）两类（Pinker，1997）。如果是问题，就有可能找到答案，并且即使我们还没有找到答案，我们可以预计会在未来找到。而对谜团来说，我们就无法预想会找到答案。用平柯的这些词语，我们可以把科学看作是把谜团转变成为问题的过程。事实上，平柯（1997）在写其《思维是如何运作的》一书中，就曾表达了这个意思：“许多有关思维的谜团，不管是关于脑

成像还是恋爱，都在最近由谜团升级成为问题了”（p.Ⅸ）。

心理学和世俗智慧：“常识”的问题

每个人都有一套有关如何处理我们自己的人际交往行为，以及怎样看我们自己及他人的内隐理论。社会、人格和认知心理学家也都研究过个体的这些内隐理论是些什么。研究发现，人们很难清晰、逻辑地把这些理论表达出来。事实上，人们只有在特意关注它们或发展它们不被证实时，才能意识到它们的存在。实际上，大多数人是不会遵循一套成型的理论来行事的。反而，当我们感到要对行为做出解释和说明时，往往会引用一些老生常谈的谚语、或一篓子的箴言。但是这些关于行为的“常识”本身往往是相互矛盾的，也因此是不可证伪的（有关可证伪的原则将在第二章中讨论）。

人们常常用相互矛盾的谚语，在不同时间来解释同类事件。例如，“三思而后行”是一个强调行动要谨慎的好俗语，但是我们也常听说“机不可失，时不再来”来劝告人们要行动果断。“小别胜新婚”表达了人们的一种情绪反应，“眼不见心不烦”不也同样是吗？类似的、呈对立状态的谚语还有“欲速则不达”和“兵贵神速”；“三个臭皮匠顶个诸葛亮”和“三个和尚没水吃”；“安全第一”和“不入虎穴，焉得虎子”；“异性相吸”和“物以类聚”；还有我经常告诫学生“今天事情今天做”，可是当我又和他们说“车到山前必有路”时，真希望自己没有说过先前的那句话。

这些大受欢迎的陈腔老调构成了我们对行为的一大堆内隐

的“解释”。无论发生了什么情况，其中之一都可以被抬出来解释一番。这就难怪我们都认为自己是判断人类行为或个性的高手。从适用性的角度来说，这些俗语简直是无所不能的。心理学家卡尔·特根（Karl Teigen）曾让被试评价类似上述意义相互矛盾的谚语，结果发现人们确实倾向于同时认为两个相互矛盾的谚语都是对的。特根总结出一个有趣的发现：“人们判定某个谚语的正误，可以和其判断一个与其对立谚语的正误完全无关”（Teigen，1986，p.47）。

因此，有时我们对一些内隐的心理理论是很难驳倒的，但是，正如我们将在第二章中会讲到的，正是因为它们是很难驳倒的，往往也没什么大用。更何况，即使其中的某些世俗信念是可以被验证，它们往往被心理学研究证实是错的。例如，对高中生课余打工一事，大多数美国人都认为是件好事，理由如下：①打工挣到的钱可以贴补他们未来的教育费用和家庭开销；②可以发展他们的工作情操，使他们在日后的职业生涯中可以承担更多的责任；③可以帮助他们形成对美国经济的正确看法；④因为已经融入了经济体系，会提升他们的学习积极性。

发展心理学家们曾作了大量有关打工对学生的影响的研究，涉及了对学生的行为、态度和学业成绩等多个层面（Bachman & Schulenberg，1993；Steinberg，Brown，& Dornbusch，1996；Steinberg，Fegley，& Dornbusch，1993）。他们发现，美国文化中对青少年打工的所有世俗看法都是错误的。事实上，只有很少的打工收入真正贴补了家庭的开销或作为未来的教育支出。大多的收入都花费在拿来比

阔的奢侈品或广告鼓吹的“必需品”上。而且，高中阶段的打工经历普遍对学生不具有教育和提升经验的意义。更有趣的是，令人感到讽刺的是，打工经验反倒让学生对工作以及其对我们经济所可能带来的贡献，变得更反感及不尊重。在研究中，打工的青少年比没打过工的人更多的抱有如：“拼命工作的人比不拼命的人更有些精神不正常”和“天下没有公司会真正关心其员工这回事”之类的想法（Greenberger & Steinberg，1986）。格利伯格和斯滕博格在回顾了大量的同类研究后，得出结论“打工绝对是弊大于利的，甚至会引发青少年的一些失足行为”（p.6）。很显然，我们给年轻人描绘了一个关于工作的虚幻图景。那些认为打工可以帮助青少年“塑造性格”和“形成对金钱的观念”等文化口号，其实都是错的。因此，我们可以说，这些老生常谈就类似于人类学家在研究发展中国家时喜欢挖掘的古老传说一样，它们令我们自己感觉舒服些，也令我们当下的文化行为变得合理，但是却毫无现实基础。

我们还可以很容易的举出许多其他的实例来说明世俗常识或民俗谚语的谬误。例如，一般公众认为在学术上成绩出色或经常读书的小孩，通常在社交及体能方面表现不好。虽然这是完全错误的认识，但它仍然在社会中广泛流传。大量事实表明，这类小孩体能优秀，在社交上也更加积极（Gage & Berliner，1984，pp. 18 ~ 19；Market Facts，1984；Zill & Winglee，1990）。例如，在学校中成绩良好的孩子比成绩不良的孩子更容易赢得朋友的接纳（Gage & Berliner，1984）。经常读书的人比不常读书的人更喜欢做运动、慢跑、露营、

踏青、修理汽车等（Zill & Winglee，1990）。

许多世俗观念一旦出现，就长久流行于世。例如，在20世纪80、90年代有一种世俗观念认为：低自尊直接导致了攻击行为。但是，实征研究显示攻击行为和低自尊之间没有联系。如果有的话，反倒是：攻击行为和高自尊经常联系在一起（Baumeister，1999；Baumeister，Boden，& Smart，1996）。另一个在20世纪90年代，极其流行的一个假设是，低自尊导致学业成绩不佳。后来证明，自尊与学业成绩之间的关系似乎与这一广为学校管理者认同的假设正好相反，是优异的学业成绩，或其他方面的成就导致了高自尊（Kahne，1996；Ruggiero，2000；Stout，2000）。

拉德福德（Radford）曾谈到过一个似乎无法泯灭的世俗迷信，即“我们仅仅用了我们脑力的10%”，这个观念尽管没有神经科学的根据（见Beyerstein，1999；Higbee & Clay，1998），但这种认识已经流传几十年了，而且好像已经获得了是“心理事实”的地位。拉德福德引述专栏作家罗勃特·塞缪森（Robert Samuelson）给“心理事实”下的定义“虽然没有坚实的事实证据，但却由于在日常生活中的不断重复，而被认为是真实的观念”（p.53）。其实，这就是世俗信念。

世俗的信念并不是完全可以不靠事实证据来持续的。有时，一旦反证被广为人知，世俗信念也会改变的。例如，多年前流传的一条谚语是“早熟早烂”（Fancher，1985，p.141），认为儿童早熟将意味着日后生活的毁灭。但是现在，心理学研究表明这是个错误观念的事实已在公众文化中生根，也就再也听不到这一谬论了。

这个例子也警示我们小心对待当今的“常识”，因为要让昨天认为正确的常识在今天变为谬论并不是难事。毕竟，“常识”是指尽人皆知的知识。那么，就像过去150年来，在美国曾经是每个人都知道的“常识”，如“女人不应该有投票权”，“美国黑人不应该接受教育”，或“残疾人都应该送进收容所”等，现在不都被我们否定为是垃圾了吗？主要就是因为它们都没有实证研究的支持。从这些例子，可以看到，相对于常识而言，心理学的用武之处，在于它以实验为基础，验证这些常识背后的假设。有时，这些关于常识的假设并不能通过检验，从而使充当检验者角色的心理学难免与文化信念发生冲突。心理学也因而成了勇敢的揭发者，让这些早为人们所惯性接受的世俗智慧再无立足之地。这就不奇怪，为什么很多人不但不信心理学的检验报告，也对做出这些报告的心理学家视若无睹。

作为一门新兴科学的心理学

在实证心理学的发展过程中，反对的声音总是不绝于耳。在100年前，英国剑桥大学拒绝建立心理物理实验室，因为这类研究会“将人的灵魂放在度量尺上，从而侮辱了宗教的神圣”（Hearst，1979，p.7）。心理学家保罗·梅尔（Paul Meehl）曾经提过，“在当代社会科学所使用的实验和统计方法尚未发展起来之时，对于人们只靠‘尽人皆知’所获取的知识（这里的人是指只要是人类的一员即可），几乎从来没有进行过任何批判性的检验”（Meehl，1991，p.440）。基

于此，心理学欲建构可实征之课题的战争可谓刚刚才取得了胜利。但是随着科学的进展，许多问题逐渐变成可以用实征方法来解决，心理学家将去挑战人们更多的固有观念。诸如，道德推理的发展、恋爱心理学、种族偏见的本质、宗教信仰的心理及社会决定因素等，都是现在心理学家正在探索的问题。最近，关于幼童性行为的研究引发了很大争议（Hagen，2001；Rind，Tromovitch，& Bauserman，2001）。有些人仍然反对对这一领域进行实征研究（Hunt，1999），但是心理学在这一领域还是取得了相当的进展。

莱维和多恩尔曾总结出，一些人反对心理学研究的原因在于他们宣称他们“没有需要知道”（Levin & O’Donnell，2000）。他们叙述，当一个校董会提出要家长来决定是否要他们的孩子在有不同年龄学生的班级学习时，有教师提出应该先做一个研究看看再说，但学校董事会否决了这一建议，因为他们认为不管“多龄”教学或“单龄”教学，哪种方法更有效，家长都会强迫学校完全转向那一较好的教学方式。莱维和多恩尔（2000）从而提到：“（对于可能的研究结果）校董事会根本就不想知道！”（p.66）。因此，我们应该充分意识到人们常常否定心理学研究，并不是因为他们觉得它不好，而是因为他们想要逃避由此所可能带来的麻烦。

也许你已经意识到了，对心理学的批评中有一个很大的矛盾之处。一方面，一些人反对把心理学称为科学，否认心理学家可以经实征来验证有关行为的事实。另一方面，另一些人由于害怕心理学家有可能发现某些行为领域的事实，而反对心理学家在这些领域里做研究。斯金纳行为学派的心理学家总

是必须面对这类矛盾的批评。例如，反对者说行为主义的强化律不适用于人类行为。同时，另一些反对者又害怕人们会运用这些规律去规范及控制他人的行为。这样，行为主义者腹背受敌，一些批评者否认行为主义者所找到的行为规律是可以应用的，而另一些批评者则害怕这些规律被滥用。

以上这些现象的产生主要是由于心理学是一门新兴的科学，刚刚开始探索以前人们没有想到要研究的行为问题，它的青涩多多少少也解释了为什么它自身作为一门实征科学的地位一直受人质疑的原因。在第十二章中，我们也将讨论到心理学自身在其发展过程中的所作所为，也对外界的这种误解产生了推波助澜的作用（Dawda & Martin，2001）。这些行为忽略了心理学是一个依靠实征研究方法的学科，从而模糊了心理学与其他领域的重要区别。由于心理学起步较晚，许多研究领域现在还只是在婴儿期，刚开始步入研究成果收获期，以及建立有解释力的理论的阶段。但无论如何，在过去的40年里，心理学已经稳固地建立在我们称之为科学、其中各学科紧密联结的知识结构之中。

小　结

心理学是一个主题非常广泛，但相对松散的学科。它包括一些不能用同一个内容来概括的众多研究主题。然而它可以通过“都是用科学方法来研究行为”这一标准被整合为一门学科。科学方法绝非是指一套生硬的规则，而是指一些非常通用的原则。最重要的三点是：①科学采用了系统实征主义

的研究方法；②它是以可以公开予以验证的知识为研究目标；③它努力寻求用实征方法可以解决的，而且可以发展出可以付诸验证之理论的研究课题（下一章的主要内容）。系统实征主义所指的有结构的及有控制的观察是本书随后几个章节的主题。科学通过同行评审的程序和重复验证的机制来保证知识的公开性。

心理学是一门新兴的科学，因而经常会和世俗的智慧相冲突。这种冲突是任何新兴科学都会遇到的，了解这个冲突有助于我们理解那些对心理学作为一门学科所带的敌意。但是与世俗常识之间的冲撞同时也令心理学成为一个非常让人振奋的学科。现在很多人加入了这个领域，正是因为它可以提供机会来检验那些存在了几百年的、被广泛认可的“常识智慧”。

（程乐华　译）

第二章　可证伪性

——如何阻挡脑袋中的鬼精灵到处捣蛋

1793 年，美国内战爆发期间，费城曾流行一种严重的传染病——黄热病。当时费城的一个名医，也是美国《独立宣言》的签署者之一——本杰明 · 拉什（Benjamin Rush），据说是少数几个可以治愈黄热病的内科医生。拉什的药物理论认为，凡伴有高热症状的疾病都必须通过大量放血（用手术或水蛭虫吸走体内的血液）来医治。他用这种方法治愈了很多病人，其中还包括他自己。然而，有人却指责他的放血疗法比黄热病本身更令人恐怖。黄热病流行过后，拉什更加坚信放血疗法的疗效，虽然事实上已经有好几个接受放血的病人在治疗中死去。他是怎么得到这一结论的呢?

有人这样描述拉什的观点："一方面坚信自己的药物理论是正确的，另一方面却缺乏有效的方法对治疗效果进行系统研究来证实自己的观点。因此，他把每一个病情好转的病例都归结为治疗方法的功效，而把每一个失败的病例都归结为病情本身的严重性"（Eisenberg，1977，p.1106）。换句话说，如果病人情况好转，那么这种好转就被看作是放血疗法有效的证据；如果病人去世了，那就表明病情已经严重到任何治疗方法都无法奏效的地步。现在我们知道对拉什的这一批评是正确的：他的放血疗法的确和黄热病同样危险。在本章里，我

们将讨论拉什究竟错在哪里。他的错误在于违反了科学思维的一个重要原则，这一原则同时也是评价心理学理论特别有用的一项标准。

本章还将详细探讨在第一章中曾提到的一个话题，关于科学的第三个普遍特征：科学家只能研究“可解决的问题”。科学家所说的“可解决的问题”通常意指“可验证的理论”，他们通过保证理论在实征观察时，有可以被验证是错的可能性（即“可证伪性”），来确保其可验证性。也就是说，一个理论所欲解释的行为现象一定要在现实生活中能表现出来，以致可以通过观察事实或现象是否存在，得知该理论是否是对的。我们将在下面看到为什么可证伪性这一标准在心理学中是这么重要。

理论与可证伪性标准

拉什在评估放血疗法的疗效时掉进了一个致命的陷阱。他评价证据的方法使他不可能得出该疗法无效的结论，因为如果病人的康复是对疗法的肯定（从而也是对他的医疗理论的肯定），那么病人死亡就应该意味着是对其疗法的否定，这样才算合理。但是，他却把这种否定合理化掉了。拉什这种解释逻辑违反了建构和检验科学理论的一条重要原则：令他的理论永远不可能被验证是错的。

在表述科学理论时，必须遵循这样的一个原则：从理论推导出的各种预测要有被证伪的可能性。在评价某一理论的新证据时，必须看在收集该新证据时，是否有可能也证实该理论是

错的，这个原则就是可证伪性标准。著作在科学工作者中间广为流传的科学哲学家卡尔 · 波普尔（Karl Popper）就一直竭力强调可证伪标准在科学进程中的重要作用（Magee，1985）。

可证伪性标准是：要说一个理论是有用的，那么根据它所作的预测必须是具体的。也就是说，理论在告诉我们如果它是对的话，哪些事情应该发生的同时，还必须说清楚哪些事情不会发生。一旦后者发生了，我们就可以清楚地知道：这个理论有些地方出错了。这时就有必要对该理论进行修改，或者提出一个全新的理论来取代原来的理论。这两种做法都会使我们提出的理论更接近真实情况。相反地，如果一个理论的预测包括了所有可能观察到的数据，那么这个理论将永远不需要也不会修改，我们也就将停滞于现有的思维槽臼而不会取得任何进步。因此，一个成功的理论并不是那种可以解释或预测任何一个可能出现的事实或现象的理论，因为这样反而恰恰丧失了该理论的预测能力。

由于本书其余章节经常对理论进行评估，所以有必要先澄清一下这个经常被误解的名词：理论。人们挂在嘴边的一句口头禅，就是"噢！这仅仅是一个理论罢了"。从这句话不难听出，外行人在使用"理论"一词时，他们的意思是指一个未被证实的假设，仅只是一种猜测，一种预感。但这决不是科学意义上的"理论"。科学家的"理论"，不是指没有被证实的假设。

科学的"理论"是指一套具有内在联系的概念，这套概念可以解释一组已有的数据（所反映的事实及现象），并且

能对未来尚未进行的实验结果进行预测。"假设"则是指由理论（相对假设而言具有较概括和较广泛的意义）所推衍出来的更为具体的预测。现存比较好的理论都是那些由其所推衍出来的假设已经得到验证的。也就是说，这些理论的结构到目前为止是与众多的实征观察结果相吻合的。但是，一旦观察数据开始与理论所推衍出来的假设产生矛盾时，科学家们就会重新去建构另一个新的理论，以便给所观察到的数据提供更好的解释。因此目前正在进行科学讨论的理论，都是在某种程度上已被实征验证过，它们所作的预测到目前为止还没有和观察到的数据产生矛盾的理论，因此它们不仅仅只是猜测或预感而已。

外行人和科学家使用"理论"这一术语的差异，经常会被一些虔诚的正统基督教徒所利用，用之作为将上帝创世说带入公立学校之讲堂的支持证据（Pennock，1999）。他们的辩词通常是"毕竟人类是由物种进化而来的进化论仅仅是一个理论而已"*。这种论调中所用的"理论"，就是外行人所理解的"仅仅是一种猜测"的意思。然而，主张"自然淘汰"的进化论并不是外行人所说的"理论"（而应是他们所说的"事实"），而是科学意义上的"理论"，它是有大量观察数据支持的构念体系（Berra，1990；Maynard Smith，1998；Ridley，1996、1999）。它不等同于那种"纯属猜测"的说

* 在美国除了私立的带有宗教色彩的学校之外，由政府资助的公立学校规定一律不准在上课时宣传"上帝创造宇宙"的说法。理由是进化论已经指出人类是物种进化而来的。反对这项规定的人则辩称进化论只是一个理论而已，而"人是上帝创造的"这一说法也是一个理论，为什么就不能在公立学校的课堂上讨论呢？

法。著名生理学家西欧都萨斯 · 杜赞斯基（Theodosius Dobzhansky）曾写了一篇很有名的文章，讲到这一点，他的论文的题目就是“生理学中除了进化论以外，别无他物”（Dobzhansky，1973），说明这一理论的重要性，不能等闲视之。

敲门节奏理论

下面举一个例子说明可证伪性原则是如何起作用的。我的一位同事认为不同类型的人，敲门节奏会不同，他说他可以据此来推测来访者为何许人物。例如，有一天当一位学生来敲我办公室的门时，在我开门之前，这位同事预测敲门者是一位女性。我打开门，果真是一名女学生。我对该同事说，他的理论很难说服我，因为即使不根据他的“敲门节奏理论”推衍，他也有 50% 的几率猜对。同事听了之后说他的理论可以有大于 50% 的预测率。又来了一个人敲门，这一次该同事告诉我是一位 22 岁以下的男性。我打开门，是一位我认识的男学生，他刚刚高中毕业，小于 22 岁。于是我告诉该同事，这一次他的理论有一点说服力了，因为我们学校 22 岁以上的学生有很多，22 岁以下的确很少，因此如果只是靠瞎蒙，猜错的几率较高。不过我仍然坚持这一次还是太容易猜对了，因为年轻的男性在校园里本来就很常见。该同事看我很难侍候，于是提出再做一次检验。当第三个人来敲门时，他预测说：“女性，30 岁，1.57 米高，左手拿着一本书和一个钱包，右手敲门。”我打开门，发现他的预测完全正确。这一次，我的反应绝然不同了。如果这 3 个人不是与我同事预先串通好的话，那么现在他

的理论可以说完全说服了我。

为什么我会有如此不同的反应呢？从一开始的"那又怎样？"，到后来的"哇！"。答案就在于预测的具体性和精密性有所不同。越具体的预测在得到证实后，越能给人留下深刻的印象。但是请注意，预测的具体性是直接与其证伪性挂钩的。越具体、越精密的预测，就越可能被将来观察到的数据所推翻。比如，有太多的人并不是 30 岁、1.57 米高的女性。我对我同事之预测的不同反应，隐含着这样的意思：一个理论的预测越是能够把依其理论认为是不可能发生的事情之范围扩大，该理论就越能给我留下深刻印象。

好的理论所作的预测总是显示自己是可以被证伪的。差的理论则不会置自身于被证伪的危险中，它们所作的预测通常非常概括，以至于总是会被验证为正确的（比如，下一个敲门的人年龄在 100 岁以下），或者它们是用完全不可被证伪的语句来表述（像拉什的例子那样："治愈就是理论正确，不愈就是病人已经病人膏肓"）。事实上，一个理论如果被置于"不可被证伪"的保护下，它便不再是科学。这也正说明为什么波普尔是在试图划分科学与非科学的差别时，提出了可证伪性原则的重要性。这一原则与心理学关系密切，也与第一章中有关弗洛伊德的讨论直接相关。

弗洛伊德与可证伪性*

上世纪初的几十年里，波普尔一直在寻找，为什么有一些理论使科学知识停滞不前，而另一些却导致知识的长足进步（Hacohen，2000）。例如，爱因斯坦的相对论。它之所以能够导致后面一系列惊人的发现（例如，从远距离恒星发出的光线在行进太阳时变弯），正是由于这一理论的预测结构与很多有可能发生的事件或现象相矛盾，因此其理论是很有可以被证伪的。

波普尔指出，一些使知识停滞的理论却并非如此。例如弗洛伊德的精神分析理论，就是用复杂的概念结构对人类行为进行事后解释，而不做事前预测。波普尔认为，正是因为它可以解释一切已发生的事情，所以该理论在科学上毫无用处，因为它不作具体的预测。精神分析学派的追随者花费大量时间和心血使该理论能够解释所有已知的人类事件，从个人的怪癖到广泛的社会现象。但是他们令该理论成功地成为事后解释的实厚依据，却使其丧失了任何科学价值。与其把弗洛伊德的精神分析看作当代心理学的一个理论，还不如把它当作是文学创作的灵感，更为恰当一点（Robins & Craik，1994；Robins，Gosling，& Craik，1999、2000）。它在心理学领

* 译者注：最近已有大众媒体（Newsweek，2002年，11月11日，46-47页）报道，弗洛伊德理论已经在认知神经科学领域，用脑成像技术得到支持。这可以说是弗氏理论，多年受正统科学心理学排斥以来，一个“扬眉吐气”的突破。这也说明，科学的方法及技术是日新月异的，眼前没法得到支持的理论，并不一定就是错的。但是不管科学的方法及技术如何的进步，其科学验证的逻辑及程序是不变的。

域的消亡，部分原因就是它不能满足可证伪性这一标准。

正如一位知名评论家指出的那样，“不正确但被广泛传播的有关心理学的想法，无可避免地会对整个社会带来灾难。由于精神分析学派的声望一度被人为地抬高，令社会上许多有疾病及遗传缺陷的人延误了有效的治疗，转而从个体早期经历中寻找自己现有痛苦的根源”（Crews，1993，p.65）。再以抽动秽语综合征（Gilles de la Tourette syndrome）的治疗史为例，这种病的症状是发病时躯体的某一个部位会发生抽搐和痉挛，并伴有言语症状，例如发出咕哝声和狗吠声，模仿言语（echolalia，无意识地重复他人的语句），秽语癖（coprolalia，强迫性重复淫秽语句）等。这是一种器官型的中央神经系统失调症，现在已经可以使用药物氟哌啶醇（haloperidol）根治（Bower，1990、1996a）。然而在它的治疗发展史上，抽动秽语综合征患者却饱受过种种残酷的迫害。早期他们被宗教权威视为是妖魔而被迫害，后来又要被迫接受“驱邪”的仪式（Hines，1988）。更重要的是，在1921～1955年之间，精神分析学派的概念占据了对该病症之病因及治疗研究的主导地位，许多学者用无法证伪的精神分析解释来理解该病，而这些似是而非的解释模糊了人们对该症本质的认识，从而阻碍了对其真正病因及治疗的研究（见Kushner，1999）。例如，有一位学者曾经这样说：

抽动秽语综合征是精神分析学派理论对脑疾病研究产生消极影响的一个典型例子。最初人们认为抽动秽语这种病是由脑的退化过程引起的。但20世纪初期，当弗洛伊德的理论成为时尚时，人们开始把注意力偏离了脑……其结果是使患该类疾

病的病人转而求救于使用精神分析法的精神科医生，而不去求治于神经科的大夫，从而忽略了对身体器官的检查和研究。（Thornton，1986，p.210）

夏皮罗（Shapiro）等人曾提到当时许多精神分析师对抽动秽语综合征的一些现在看来很可笑的想法（Shapiro et al.，1978）。其中一位认为他的病人“不愿意放弃抽搐，是因为对她而言，这种抽搐已经变成性快感的源泉和潜意识性欲的表达方式”。另一位精神分析师则认为这种抽搐“已经等同于手淫……与生殖器相联系的里比多（libido）错放在了身体的其他部位”。第三个把这种抽搐看作是“肛门－被虐转化期的症状”。第四个认为抽动秽语综合征患者具有“强迫性人格和自恋倾向”，病人的抽搐“代表或者曾经代表一种情感的症状，是对想要表达情感的压抑性防御”。夏皮罗等人（1978）下面的一段总结很好地说明了忽视可证伪性这一标准所可能会对理论带来多坏的影响：

精神分析学派建构理论的方式实际上面面俱到。抽搐是一种转化期症状，表现为歇斯底里但又反映了性欲，受意志控制而又是强迫性的，是器官性病变，同时又与心理冲突有关……所有这些心理学的标签、诊断和治疗，很不幸地通通都贴到了病人及其家人的身上。而且这样做的过程中，治疗者对待病人及其家属毫不客气，相当武断，从而给他们造成很大的伤害……由于这些论文后来产生了广泛影响，因而给人们认识和治疗这个疾病造成了极大的障碍（pp.39～42，50，63）。

后来，当研究人员认识到，精神分析理论的“解释”对

治疗该疾病毫无用处时，进一步认识和治疗抽动秽语综合征才变为可能。但那些毫无用处的解释往往相当诱人，因为它们看似在解释事情。但是事实上它们确是解释了所有事情，不过都是在事后而已。精神分析理论所给予的解释只创造了理解的假象而已。由于要努力解释所有事后发生的事情，所以这一理论本身得不到进展。一个理论只有在不去预测任何事件都可能发生，而只是具体地去预测，即事前告诉我们什么特别的行为会出现时，进步才会发生。当然，由这类理论所推导出的预测很有可能被证实是错误的，但这正是科学研究的长处，而绝非其缺点。

会隐形的鬼精灵

如果人们能够从所研究的问题抽离出来一下的话，就不难察觉哪些理论的构想是不可以证伪的。特别是如果人们能汲取前人错误的教训的话（就像拉什的例子那样），那就会更容易察觉了。另外，当事例很明显是编造出来的时候，也很容易察觉出其不具可证伪性。例如，我发现了人的大脑中有一个可以控制行为的潜在机制，一开始这一发现还鲜为人知。但不久之后，你就会在一些随处可购得的八卦报刊上读到有关这一发现的报道。这一潜在机制是这样的：在大脑的左半球靠近言语区域，住着两个鬼精灵。他们有能力控制发生在大脑多个区域的化学放电过程。也就是说，这两个鬼精灵基本上是控制了一切。不过有一点，这两个鬼精灵很狡猾，有本事觉察一切对大脑的探测活动（无论是通过手术或是X线之类的途径），一旦他们感觉到外界的探测，他们就会悄悄地躲

藏起来（哦，我刚才忘记说，他们有隐形的本事），所以人们永远无法证实确有他们的存在。

你也许认为我用这样一个骗小学生的故事来做例子，是侮辱了你的智慧。但是如果你能设想，你像我一样作为一个心理学导论的授课老师，却经常被学生问及为什么我在课堂上不和大家讲有关超感知现象（extrasensory perception，ESP）这一领域的惊人发现，或是为什么不讲过去几年在通灵学所取得的进展的话，也许你就不会怪我为什么会用这么幼稚的例子来说明问题了。回答这些学生所提出来的这些幼稚的问题，我不得不告诉他们：听到的大部分有关这些方面的信息，无疑都来自大众媒体，不是来自有学术价值的科学刊物。事实上，有一些科学家已经在调查及重验这些媒体炒作的研究成果，但是都不成功。我进而提醒这些还没有完成心理学研究方法课程的学生，一个研究成果能否成为一个科学事实，并被人们所接受，是取决于它的可重验性，特别是在我们遇到研究结果与过去已收集到的数据或者现有理论所预期的相矛盾的时候，这一点尤为重要。

让我进一步指出，很多科学家现在对有关超感知现象的研究已经到了忍无可忍的地步，原因之一固然是该领域已经被诈骗、江湖郎中、媒体剥削所搞臭了。但是更重要的却是，科学家们已经察觉到，如马丁·加德纳（Martin Gardner）曾经指出的，超感知现象研究的无法被证伪性（Gardner，1972）。

超感知现象的研究通常是这样进行的：首先，“相信者”（即在研究开始前就已经相信超感知现象存在的人）宣称已经在实验室里证明了超感知现象的存在。然后，一位“怀疑者”

（对超感知现象的存在持怀疑态度的人）被邀请到实验室去证实超感知现象的存在。通常怀疑者在检查完实验的内外环境之后，都会建议"相信者"对实验条件进行更严密的控制（关于实验控制的类型，将在第六章详细讨论），以减少"相信者"从中做假的可能性。虽然有时这些建议会遭到拒绝，但大部分善意的相信者是会"从善如流"的。一旦这些较严格的控制条件都设置好了之后，超感知现象就不复存在了（见 Alcock，1990；Druckman & Swets，1988；Hooft，2000；Humphrey，1996；Hyman，1992、1996；Marks，2001；Milton & Wiseman，1999；Wiseman，Beloff，& Morris，1996）。然后，怀疑者就会说：实验的失败说明，原来的实验是因为缺乏有效的控制，才会得出超感知现象存在这一错误的结论。但他们常常会惊讶地发现相信者并不这样认为。相反地，这些相信者会用不可被证伪的方式来解释超感知现象：他们坚持认为心灵的力量，例如超感知是非常隐晦的、细致的，而且很容易因被干扰而失灵的。怀疑者"怀疑的颤波"很可能是这一心灵力量瓦解的罪魁祸首。当怀疑者反对的"气氛"被移开时，这种心灵的力量肯定会再回来的。

令人惊讶的是，在听完我解说了相信者的心态之后，学生们不但没有失笑，反而恍然大悟的说，"呀！这很对呀，我说嘛，原来是怀疑者所发出的负面干扰把超感知现象给搞没了。"听到学生们的反应，我虽稍感惊讶，但仍然保持镇定，并耐着性子提醒他们想想在几个星期前的课堂上，我举的会隐形的鬼精灵的例子。我说，当时全班同学都认为鬼精灵的存在证据，会因为一要探究它就隐身不见了，这一说法

是非常荒谬的，现在相信者提出的超感知现象的存在证据与会隐形的鬼精灵具有同样的性质。超感知现象运作的方式与鬼精灵是一样的——只要你不闯进去仔细地观察它，它就好好的在那里。但是，你若闯进去想看个究竟，它就马上隐形消失。如果我们接受这种解释，那么就永远也无法证明给怀疑者看到它的存在，因为照相信者的解释任何怀疑者的出现都是令其失灵的原因。它存在的证据只能在相信者面前显现。显然，这种说法是不能被科学所接受的。在科学的领域里，我们没有磁力学家和非磁力学家（亦即那些相信磁力存在和不相信磁力存在的物理学家）之分。在这一点上，学生显得有点胆怯，他们咕哝着说同意我的观点后，迅速离开了我的办公室。当然，我永远不知道学生的这种同意是意味着真正理解，还是意味着尴尬因此想尽快逃离现场。也许，两者兼而有之，因为对学生来说，他们正面临一件世界上最困难情景：面对与自己所坚信的正相反的事实。严谨地运用科学准则来验证自己长期固守的信念是相当令人不舒服的事，但是这是一个心理学家必须做的。从来没有人说这会是很容易做到的。

并非所有的验证都是一样的

可证伪性原则对我们要如何看待理论验证具有重大意义及影响。很多人认为一个好的科学理论是一个被反复验证了的理论。他们认为被验证的次数对评价一个理论非常关键。但是前面对一个理论的可证伪性的讨论，让我们看到理论被验证的次数并不是关键。因为正如前面所举的“敲门节奏理论”例子那样，并非所有的验证都是一样的。验证能否令人信服，

取决于理论所推衍出来的预测本身可否能被证伪。验证 20 个实际上不可能被证伪的预测（例如，小于 100 岁的人），远远不如验证 1 个非常具体、有可能被证伪的预测（例如，女性，30 岁，1.57 米高，左手拿着书和钱包，右手敲门），更能令人信服。

看来，我们不应只注重理论被证实的次数，还应该注重验证本身的质量。采用可证伪性作为评价标准，可以帮助使用研究成果的人抵制伪科学最喜欢声称的“全能”（可以解释一切的理论）的诱惑。这些全能理论不可避免地会妨碍我们对世界和人类本质的进一步认识。这些理论死巷，无可否认地，有时相当具有吸引力，正是因为它们无法被证伪。它们在现代这个动荡混乱的世界里，常被视为是稳定的安心岛。

波普尔经常谈到“这些（不可证伪）理论之所以会在心理上对人们产生这么大的吸引力，其奥秘就在于它们能够解释一切。知道无论发生什么事，你事先都可以找到解释它的答案，不仅让你觉得你掌握了知识，更重要的，它还给了你应对这个世界所需的安全感”（Magge，1985，p.43）。然而，获得这种安全感并不是科学研究的目的，因为它是以知识发展的停滞为代价所换取得来的。科学是一个不断挑战先前存有观念的机制，通过实征研究来设法证明前有理论的缺陷。这一特点常常使科学，尤其是心理学，与所谓的世俗智慧或一般常识发生冲突。

可证伪性与世俗智慧

心理学可以说是威胁了人们运用第一章所提到之世俗智慧

的安逸感。因为作为一门科学，心理学不能安于用这种驳不倒的解释来理解人的行为。心理学就是要对各种不同有关行为的理论，逐一进行检验与筛选。当然，心理学很欢迎世俗智慧中表述清晰、可站出来接受检验的内容，而且很多这类内容也已经被整合到心理学理论当中了。只不过心理学不会去寻求由事后可以解释一切，但事前作不出任何预测的解释系统所提供的安逸感。心理学也不接受那些永不需更新，并可从一代转给另一代的世俗智慧系统。如果你我不向学生或者公众坦白承认这一事实，我们只会像是用石头砸自己的脚那样，自求毁灭。遗憾的是，有些心理学教师或主张将心理学大众化的人却往往担心，心理学对世俗智慧构成的这种威胁会困扰人们，于是就传递了一些错误的信息，来舒缓这些困扰："你将学到一些有趣的知识，不过不用担心，心理学不会挑战你深信不疑的东西。"这种说法是错误的。因为会令人们对什么是科学，以及什么是心理学更加疑惑。

科学寻求的是在想法上的改变。科学家是在努力描绘世界的真实面目，而不是去描绘我们固有的信念所认为应该有的世界之面貌。在现代思潮中有一股危险的趋势认为：我们要尽量避免让一般大众知道事实的真相——在这些人的面前遮上一层无知的面纱是必要的，因为他们不曾学会如何处理事实真相。这样的观点，心理学和其他科学一样都会予以驳斥。生物学家米歇尔 · 吉瑟林（Michael Ghiselin）曾进一步指出，如果知识得不到传播，我们全都会蒙受损失：

如果我们的邻居都很健康，那么我们也会更好。为了要比别人更健康而垄断所有的药物供应，不给周围的人用，是

非常愚蠢的行径。对待知识也同样。邻居的无知就如他们的身体不健康一样，这对我们自己也是不好的，而且还可能是造成我们不好的原因。工业生产及我们所有人都受益于有熟练的劳动力。我们都是要依赖于其他人的技能和特长的（Ghiselin，1989，p.192）。

当我们周围充斥着对人类行为有误解的人时，心理学家也会像吉瑟林那样，感到我们都蒙受损失。世界是由公众对教育、犯罪、健康、工业生产、儿童福利，以及其他很多关键问题的众多态度所塑造而成的。如果这些态度都来自不正确的行为理论，那么，我们所有人都会受到伤害。

允许“承认错误”

可证伪性原则最有用、最具解放性意义之处，在于它不认为在科学上犯错误是一种罪过。被证伪的假设也能够为科学家提供信息，以使他们对自己的理论做出调整，从而使之更加符合观察得来的数据。哲学家丹尼尔·丹尼特（Daniel Dennett）说过，科学的本质在于“在公众面前犯错误”（Dennett，1955，p.380）。当收集来的数据与理论不一致时，通过不断修改理论，科学家最终能建构出一个反映世界本质的理论。可证伪性是理论内在的更新机制，我们不应像在处理日常事务那样去避免这些错误。

事实上，如果我们在处理个人日常生活中的事务时也使用可证伪性原则，也许我们的生活质量也会得到很大的改善。这就是前面我使用“具解放性意义”一词的原因。解放性是特别针对

其对个人的意义而言的，因为由可证伪性原则所引发出来的意义已经超越了科学意义。如果我们明白，当我们所深信的事情与所观察到的事实相悖时，最好去改变我们的信念，而不是去否认事实的话，那将会令这个世界减少很多社会和个人的问题。物理学家J · 罗勃特 · 奥本海默（J. Robert Oppenheimer）曾说过：

在任何一个人的推理思维过程中，总会有一个时候他会回顾自己的过去说:“这个问题我当时没想明白”。靠其行业过日子的人们也都需要能克服自己的羞耻感，勇敢地承认，“我的看法是错的”的能力。科学研究是让这个认错的过程每时每刻都可能发生。往往你会在思考了很长时间的一个问题上，发现了矛盾及不寻常的现象。由这些现象所带来的震惊，可能会促使你去重新思考。这与有些自认为见识很广的人不断地试图去寻找一些合理说法，来为自己最初所犯下的错误辩解相比，可谓是大相径庭（Dos Passos，1964，pp.150～151）。

请试想一下，过去有多少次，正当你与他人激烈争论时，突然间——也许正当你刚刚给予对方一次猛烈的反击来维护自己的观点时，发现在某些关键问题或关键证据方面，你的看法的确有错误。那时你是怎样处理这个错误的呢？你可曾向对方承认你的错误，并且肯定对方的看法现在看来的确比你的观点更合理？你或许不曾这样做。如果你和绝大多数人一样，那么你一定会“没完没了地寻找一些合理说法，来为自己的错误辩解”。你会拼命地使自己在不宣告失败的情况下从争论中摆脱出来。承认自己错了恐怕你是绝对不会做的。但是，其实它却是你惟一能做的。因为如果辩论双方在争论之

后，都仍不知：“到底谁的观点更接近真理呢？”如果争论的内容没有公布于众（不像在科学研究中那样一定要公布于众），如果双方争论的力度势均力敌，如果对争论的结果（例如，知道自己犯了错误）不能正确的反馈给对方（好像前面举的那个例子那样），那么将找不到适当的机制来使人们的信念与事实真相相吻合。这就是为什么那么多私下或公开的对话及沟通，都只产生了令人更加困惑的结果的原因。这也是为什么心理学在解释人类行为的原因方面，比起一般常识或世俗智慧，是更可靠的资料来源。

很多科学家都见证了，在科学的进程中，犯错误是正常的事，并且知道科学发展的真正危险在于人类总想避免暴露自己的错误观点，明白这两点是相当重要的。诺贝尔奖获得者米德瓦（Peter Medawar）曾写道：

虽然提出有瑕疵的假设情有可原，因为迟早它都会被比较好的假设给换掉，但这样仍会对曾经深信过它们的人们造成严重的伤害。因为那些热爱自己理论假设的科学家们，大多不愿意承认实验的检验结果证实自己是错的。有时候，他们会不让自己的理论假设接受严格的检验，而是绕过这类检验，仅仅让假设中无关痛痒的部分接受检验，不冒可能被彻底推翻的危险……对于任何一个时代的科学家，我所能给的最好的建议就是：一个假设被坚信为真的之程度，与它是否真的为真，完全是两码子事（Medawar，1979，p.39）。

许多著名的心理学家都采纳了米德瓦的建议。在一篇报道著名实验心理学家罗勃特·克若德（Robert Crowder）之职业生

涯的文章中，引述了他的一位同事马萨瑞 · 本纳加（Mahzarin Banaji）的一段话："他是我所知道的、最不维护自己理论的科学家。如果你有办法指出他的理论有漏洞，或者他的实验结果有局限和缺陷，他会非常乐意地和你一起计划如何推翻该理论"（Azar, 1999, p.18）。艾泽（1999）描述了克若德是怎样发展出一个有关记忆之一个成分的理论，这个成分叫"分类前的听觉存储器"，然后仔细地设计实验研究来证伪自己的理论。另外一个例子是进化论心理学家约翰 · 托比（John Tooby），他在讲述达尔文在科学史上做出里程碑式贡献是得益于他对科学的一些态度时，曾作过以下这样的精彩评论："达尔文能比其同时代的人走得更远，是因为他比较不为要整个宇宙来配合自己的理论预测所羁绊（Tooby, 2002, p.12）。

但是要让科学发挥作用，并不需要每一个科学家都必须具备随时要被证伪的态度。雅各布 · 布朗劳斯基（Jacob Bronowski）在他的很多著作中提到：科学揭示世界真知的独特力量并不取决于科学家们所具备的、超于常人的某些独特品质（例如，他们是完全客观的，他们在解释结果时永远不会发生偏差等），实际上这种力量来源于可以纠正会犯错的科学家们之理论的审察过程。在这个过程中，一些科学家会竭力挖掘和批评另一些科学家的错误（Bronowski, 1973, 1977）。哲学家丹尼特（2000）也表示了同样看法，他认为没有必要要求每一位科学家都表现出如克若德那样的客观性。像布朗劳斯基一样，丹尼特强调"科学家和其他人有同样的缺点及同样会犯错，不过科学家与一般人不同的是他们意识到他们的错误是来源于自己以及其所属学术团体，因此他们

会设计出一套精细的系统来约束自己，减少犯错，努力防止因自身的弱点及偏见而影响到自己的研究结果"（p.42）。

从本质上讲，我们在这里谈的是学术的诚信，以及一种任何科学家都珍惜的开放性。用普通大众的话来说，开放性就是指，为了追求真理，对一个现象的所有可能解释，都持容忍及开放的态度。但是对科学而言，开放性思维不仅仅意味着开放的态度，它还有更深一层含义。哲学家乔纳森·阿德勒（Jonathan Adler）这样教导我们:科学更看重开放性思维的另一面，即"一个真正具有开放性思维的人是那些愿意跟随着证据走的人。这种人尊重客观公正的研究成果而不是受制于个人的偏见……科学的研究方法是与跟随世界的真相来调整自己，而不是要与我们个人的想法保持一致性"（Adler，1998，p.44）。

只有"想法"是不值钱的

从前面关于验证世俗智慧的讨论中，我们得到另一个有关可证伪性原则的小结论：想法是不值钱的。当然，我们这里的意思是说，有些类的想法是不值钱的。古生物学家、科普读物著名作者斯蒂芬·J·古尔德（Stephen J. Gould）对这个观点作了如下解释：

写了15年的每月专栏，我收到大量非专业人士写来的信件，这些信件涉及科学的方方面面……在这些信件中我发现了一个最常见的错误观念，它出现的频率远远高于别的错误观念。人们在信中会告诉我他们发现了一个革命性理论，一个可以拓展科学界限的理论。这些理论，通常要几页写得密密麻麻的纸才能陈述完毕，它们全是对我们人类最根本及最深奥的终

极问题的一些猜测——生活的本质是什么？宇宙是怎么来的？时间的起点在哪里？然而这些想法都是不值钱的，因为任何稍有智力的人都可以在吃早餐前想出几个念头来。科学家也有能力就这些终极问题提出各种想法，但是我们不这样做（或者，可以反过来说，我们尽量避免思想中出现这类空泛的念头），因为我们无法想出办法及程序来验证这些想法，从而判断它们是否正确。一个看似吸引人但却无法被证实或证伪的观点，对科学来说又有什么用呢？（Gould，1987，p.18）。

对古尔德最后一个问题的回答是："一点用也没有"。他所说的那些不值钱的想法就是前面波普尔提到的那些宏大理论。这种宏大理论是那么博泛，那么复杂模糊，以至于我们可以用它来解释一切。建构这种理论更多地是给人们提供了情感支持，因为它们的被提出并不是准备给他人去修改，或抛弃的。古尔德告诉我们这样的理论对科学的进展毫无用处，纵然有了它们会让人感到很安慰。科学是一个创造的过程，然而，这种创造只用于建构符合实征数据的理论之上。这一创造过程是很困难做到的。只有少数用以解释世界真实面目的想法不能说是不值钱的。可能这正说明为什么好的科学理论少之又少，而那些不可证伪的伪科学信念体系却遍布各处的原因所在——比起前者，后者是太容易建构了。

科学理论与世界联系紧密，它们是可证伪的，它们所作的预测是具体的。无可否认的，建构一个真正提供科学解释的理论是一项艰辛的工程，但是认识科学运作的一般逻辑却并不难。事实上，现在已有许多专门写给少年儿童看的，关于

科学逻辑思维的刊物问世（Kramer，1987）。因此科学对科学家来说虽然可能是困难的，但是从某一个角度来看，它对一般公众而言应该不是难事，只要公众愿意学习一些普遍原则——其实就只学包括在本书中介绍的这几个原则也就够了。关键是人们必须认真地掌握这些原则。好，下面就让我们开始学习吧。

科学中的错误：逐步接近真理的途径

为了解释清楚可证伪性原则，我们曾先勾勒出了一个科学进步的简单模型。理论被提出了，依据理论也推衍出了假设，然后使用一系列的实验手段（将在余下的章节中讨论）验证假设。若假设被实验证实，则理论得到一定程度的巩固；若假设被实验证伪，那么理论必须作某种程度的修改，或者被更好的理论所取代。

当然，我们说科学的知识都是暂时成立的，由理论推衍出的假设有可能是错误的，并不意味着什么人都可以来插一脚，什么想法都可以成为理论。科学里很多事物间的关系被多次证实属实之后，已经成为公理，因为它们几乎没有可能被将来的实验所推翻。比如说，人体血液不再循环或者地球不再绕太阳旋转的情况，将来不大可能出现。这些众所周知的事实不属于我们所谈的假设。它们不是科学家的兴趣所在，因为它们已经被充分证实了。科学家只对自然界那些已有一定的知识，但是还有许多未解之谜的问题感兴趣。也就是说，他们对那些还没有被证实到已经没有什么疑点的问题感兴趣。

科学实践中，科学家只看重未解决的问题而忽视已被充分证实之观点（所谓的公理）的态度，一般公众对之颇不能理解。科学家似乎总在强调他们不知道的东西而不是强调已经知道的东西。实际情况的确如此，不过科学家这样做是有充分理由的。为了获得知识，科学家必须站在已知事实的前沿，当然，前沿也正是不确定事物最多的地方。科学的进展也正是在探索如何减少这些地方的不确定性的努力过程中产生的。这就经常会令科学家常常在公众面前显得非常“不确定”，但这只是表面现象。事实上，科学家只有在知识的前沿才感到不确定，而这个知识的前沿地带正是科学即将有突破的地方。科学家是不会对那些已经被很多研究证实的事实不确定的。

这里还必须强调的是，当科学家谈到，要用实际观察来证伪一个理论，或者要用新的理论代替已被证伪的旧理论时，他们并不会完全抛弃该理论原有的那些已有实征数据支持的部分（将在第八章用大篇幅讨论这一问题）。恰恰相反，新理论必须能够解释旧理论所能解释的和不能解释的所有事实。因此理论的可证伪性并不意味着科学家非得建构一个全新的理论不可。科普读物作家爱萨克·阿瑟姆（Isaac Asimov）在他的《错误的相对性》（1989）中很好地阐释了理论的修正过程，在该书中记述了人们是怎样逐步认清地球的形状这一例子。首先，他提醒我们不要认为原来把地球看作是扁平的观点是很愚蠢的，因为在大平原上（大部分的文明社会与文字发展都起源于大平原），地球确实看起来很扁平。紧接着阿瑟姆提醒我们去想像，如果对不同的理论进行定量比较，结果会告诉我们什么？首先可以根据每米显示多少曲率度来表述

不同的理论。认为地球是平面的理论会说曲率是每米0度。现在我们当然知道这个理论是错误的。但在某种角度来看，它很接近真理。正如阿瑟姆（1989）所说：

大约在亚里士多德逝世一个世纪后，希腊哲学家埃拉托斯特尼（Eratosthenes）注意到太阳光在不同纬度投射的阴影长度不同（假设地球是平面，所有的阴影都应该一样长）。根据阴影的不同长度，他计算了地球球面的大小，结果算出地球圆周为40233.6千米，那么这个球体的曲率就达到每米0.000126度。你可以看到，0.000126与0的差别有多么细微。这一细微的差异让我们花费了好长的时间才从把地球看成平面转到发现它是球体。这里要提醒你，即使是像0和0.000126之间这样细微的差异，也是极其重要的。这个微小差异可以累积为大的差别。如果不考虑这一细微差异，如果不把地球看作球体而看作平面，那么我们就根本不可能准确地绘制出地球上大块土地的地图（pp.39～40）。

当然，科学并不满足于得到地球是圆的这样的知识。正如前面讨论的，科学家总是尽可能地提炼自己的理论，并且去挑战现有知识的界限。例如，牛顿力学理论预测地球不是一个标准的球体，这一预测后来确实被证实了。现在已经证明地球是一个赤道凸出，两极稍扁的球体，有时候也把它叫做扁球体。从北极到南极地球的直径为12713.8千米，赤道的直径为12757.3千米。地球的曲率也因此不是恒定的（不像一个正圆球体，其曲率在各处都是一样），它在不同的地方从0.202米变到0.203米不等。正如阿瑟姆（1989）所说的，

“比起从平面到球体，从标准球体到扁球体的修正要小得多。因此，虽然认为地球是标准球体的观点是不正确，但是严格地说，比起认为地球是平面的观点而言，它没有错得那么厉害”（p.41）。阿瑟姆所述关于地球形状的例子，向我们展示了科学家是在什么样的情况下，在使用像误解（mistake）、错误（error）、证伪（falsified）这样的术语。使用这样的术语并不意味着被检验的理论在每一方面都是错误的，而是说它还不完善。所以当科学家强调知识是暂时成立的，知识可能会被将来的研究结果所加以修正时，他们的意思就像这里举的这个例子所说的那样。科学家相信地球是一个球体的同时，他们已经认识到，仔细来看，这一概论是有可以被做进一步的修改的。然而，从标准球体变到扁球体，并没有完全改变地球是一个球体的正确性。所以当我们某一天睡觉醒来时，绝不会发现地球变成了一个立方体。

小　结

科学家通常所说的一个可解决的问题是指一个可验证的理论。可验证的理论在科学上的定义是非常特别的：是指可能被证伪的理论。如果一个理论不可以被证伪，那么它对真实世界毫无影响，也因此就毫无用处了。心理学界充斥着种种不可证伪理论，令其深受危害，也是造成其发展滞后的原因之一。

好的理论是指那些做出具体预测的理论，也因此它们具有很高的可证伪性。越是具体的预测，一旦得到证实，给理论提供的支持就越大。简而言之，可证伪性原则的一个重要内

涵就是：并非所有理论的验证及证实都是一样有价值的。可证伪性越高、预测越具体的理论被证实了之后，就越能被人们所接受及重视。即使预测不被证实（即它们被证伪），证伪对理论的发展也是有用的。一个被证伪的预测显示其理论依据必须被抛弃，或被依据实征数据来加以修正。如此一来，被证伪的预测还是有用的，因为它促使理论做出调整。正是通过这种调整，像心理学这样的科学才能更加接近真理。

（肖晓云　译）

第三章　操作主义及本质主义

——不过，博士，这到底是什么意思呢

物理学家真的知道地心引力是什么吗？我的意思是真正的知道地心引力一词真正的含义是什么？其内在本质又是什么呢？在提到地心引力一词时究竟是在指什么呢？如果刨根问底的话，它到底是怎么一回事？

诸如此类的问题反映了人们对科学的一种看法。哲学家波普尔称之为本质主义。本质主义认为能称之为好的科学理论必须根据现象背后的本质或内在属性对现象做出终极性的解释。持本质主义观点的人一般会认为无法对现象做出终极性解释的理论都是无用的。不能对现象做出终极性解释就无法反映内在真实情况——即世界的本来面目。本章将讨论为什么科学并不去回答这类本质主义者所提出的问题，而是借助于发展出概念的操作定义而获得进展。

为什么科学家不是本质主义者

事实上，真正的科学家从不认为自己是在追求本质主义论者孜孜以求的那类知识。对本章一开始提出的那些问题的正确回答是："就探求本质的意义上讲，物理学家的确不知道地心引力是什么。"科学家也不宣称是去求取完美无缺的知识

（科学的独特力量并非在于它是一个与错误绝缘的过程，而在于它提供一个途径以消除存在于我们知识体系中的误区）。科学也从不试图去回答关于宇宙的“终极”问题。米德瓦（1984）曾写道:

科学所不能回答的，以及在科学进展的预见范围内也不可能回答的问题，总是有的。这些问题是孩子们经常会问的……有关“终极”的问题……我所想到的这类问题包括：一切事物是从哪里开始的？我们都到这个世界来干什么？人干嘛要活着？（p.66）。

[但是]即使科学不能回答这类关于“初始”以及“终极”的问题，并不意味着它必须无条件接受对另外一些问题的答案；也不能理所当然地认为既然能提出这类“终极”问题，就一定可能找到回答这类问题的答案。以我们目前的理解力而言，这类问题是无从回答的（p.60）。

[最后]别忘了，就科学所能够解决的问题而言，科学解决这些问题的能力则是不可限量的，除了人们道德方面的缺陷——例如缺乏勇气之外，没有什么能阻碍科学知识的发展（p.86）。

因此，如果某个人，某个理论或思想体系声称能就那些“终极”问题给出绝对答案，科学家们就会对之产生怀疑，原因之一在于：科学家们认为这类“终极”问题是科学研究无法回答的。更进一步的说，断言知识的绝对性和完美性将扼杀对知识的不断探索。因为能对知识进行自由而广泛的探究是科学研究的必备先决条件，科学家们一贯质疑那些声称已经发现终极答案的说法。

本质主义者喜欢咬文嚼字

具有本质主义倾向的人通常有一个共同特点，那就是在还未开始研究术语和概念所涉及的知识前，就过分地关注如何对它们加以定义。他们常用的口号是："不过，必须先把我们要用的术语定义好。""这个理论概念真正的含义到底是什么"似乎意味着：在一个词可以被用作一个理论中的概念之前，必须对使用这个词所涉及到的各种语言问题有一个全面而且毫不含糊的认识。事实上，这恰与科学家们的工作方式相悖。在研究物理世界之前，物理学家们决不会陷入"如何使用能量一词"的争论中去；或者当谈到物质的最基本组成成分时，科学家们也不会陷入"粒子一词是否真正表达了我们所指的含义"的争论之中。

只有在广泛地研究了与某个术语相关的现象之后，而不是在此之前，才能确定相应科学概念的含义。对概念性术语之定义的修正源自在科学研究过程中所累积的、具有内部一致性的数据和其理论基础之间的相互修正，而不是源自对语词用法的争论。本质主义往往将我们带入无尽的语词争论之中，许多科学家都认为玩味这些词语游戏分散了我们对更重要问题的关注。例如，针对"生命这一词的真义是什么？"这一问题，两个生物学家的回答是："生命这两个字在我们的学术领域中没有什么直正的意义可言，它只是一个能够满足生物学家研究目的的术语而已，不是词义争论的焦点"（Medawar & Medawar，1983，pp.66～67）。简言之，科学家的目的是解释现象而不是分析语言。到目前为止，所有科学发展的关

键都在于抛弃本质主义而采纳操作主义，而这在心理学中尤为明显。这正是本章所要探讨的主题。

操作主义者将概念与可观测的事件联系起来

好，既然科学概念的意义不是得自于对语词的讨论，那么它们又是从何而来的呢？正确地使用一个科学概念的标准又是从何而来的呢？要回答这些问题，必须先讨论一下什么是操作主义，因为它对在科学研究中如何建构理论，以及对如何评估心理学中所提出的理论都至关重要。

尽管操作主义有多种形式，但对科学信息的使用者来说，以最广泛的角度来理解它就可以了。操作主义就是指科学理论的概念必须以某种形式植根于，或可联系于，可以通过测量而观察得到之事件的那种观点。将理论概念与可观测事件相联系就是这一概念的操作定义，同时也使得这个概念公开化了。这种操作性定义排除了概念中所可能涉及的、有关个别研究者情感与直觉的因素，并使得任何一个能执行这个可测量之操作的人都可以对该操作性定义进行验证。

举个例子，把饥饿定义为“我胃里阵阵难当的痛苦感觉”就不是饥饿的操作性定义，因为它与“阵阵难当的痛苦”的人之感觉经验相联系，而这一经验是不能被别的观察者所观测到的。相反地，如果定义像“未曾进食的时间”或像“血糖水平”这样的可以观测的生理指标，那它就是一个操作性定义。因为这一定义是任何人都能取得的、可观察到的测量。同样地，心理学家们不能仅将焦虑定义为“我有时会感到的、不舒服及紧张的情绪”，而是要将该概念用一系列操作来定义，

例如用问卷调查或生理测量等来定义之。前面那个非操作性定义依赖于人们对身体状况的个别解释，故而无法被其他人所重复。而操作性定义则把概念在科学领域公开化了。

认识到科学概念，是由一组的操作所定义的，而不仅由一个单一行为事件或任务来定义的，这一点是非常重要的。包括在这一组的操作；各自都在观测同一个概念，但是手段略有不同（将在第八章中详谈与此有关的趋向集中的操作）。例如，教育心理学家用在某一标准化测量工具上的表现来定义“阅读能力”这一概念，伍德库克阅读精通量表（Woodcock，1987）就是常用的一个工具。该量表总的阅读分数是由在几个分测验中所得的分数相加而来的。这几个分测验分别测量了与阅读有关的、但又有细微差别的那些技能。例如，先让被试阅读一篇文章，然后进行语词填空（找一个适当的语词填入文章内的空格中），举出一个词的同义词，正确地读出一个较难的单词，以及其他的一些测验。在这些测验上的总体表现界定了阅读能力这一概念。

操作性定义迫使我们去想清楚要如何用外界可观测的事件或任务来界定一个概念。现在，请你试想我们要给一个看来相当简单的概念——打字能力，下一个操作性定义。你之所以要这样做是因为你想要比较两种教人打字的方法。请试想一下，在完成这一比较时，你所必须作的各种选择及决定。你当然必须测量打字的速度，但是你在测量打字速度时，要学生打多长的一段文章呢？ 100个字的文章显然太短，1万字的文章当然又过长。那到底要多长才算好？打字速度要维持多久才能与研究者理论建构中的打字能力这一概念相匹配呢？要学生打哪一类的

文章才好？应不应该包括数字及数学公式，或者不常见的空格？我们要怎么处理打错字的问题？在测量打字能力时，好像应该将速度及错误率同时考虑进去才对。但是，要把这两个指标同时考虑进去的话，要如何来计出一个总分呢？是要给这两个指标同等的权重吗？还是其中一个比另外一个更重要一些？寻求一个好的操作定义会让你不得不就上述所提出来的这些问题，作一番谨慎思考，这样也会让你对打字能力这一概念本身有一个更彻底的探讨。

不同概念与其相对应的可观测操作定义之间的联系，有的比较直接，也有的比较间接，在程度上有很大差异。很少有科学概念是完全用由现实生活中直接可观测的操作来定义的。大部分的研究概念都是用与之有间接联系的操作来定义的。例如，某些概念是用一组可观测的操作和与这一个别概念有理论关联的其他概念来定义的。最后，还有一些概念不是由任何可观测的操作直接定义的，但它们是用与其他可观测概念的联系来定义的。这类概念通常被称为是隐性概念，在心理学的研究中屡见不鲜。

举例来说，由于A 型行为模式与冠心病的发病率有关（Austin & Deary，2002；Curtis & O'Keefe，2002；Dembroski & Costa，1987），所以对其作了大量的研究。我们将在第八章中详细讨论这个问题。在此仅想着重阐明，事实上，A型行为模式是由一组子概念定义出来的：具有一股强烈的竞争意识，潜在攻击性，做事火急火燎，较强的成就动机等等（Dembroski & Costa，1987）。然而每一个定义A型行为模式特征的子概念本身也是一个需要进行操作性定义的概念。事实上，许多研究者

也确实在操作定义每一个子概念上下了相当大的功夫（Dembroski & Costa，1987）。对我们现在的讨论较为重要的一点是：A 型行为模式是一个复杂的概念。它并非由操作直接定义的，相反它是由一组其他概念所定义，而这些概念则各有其操作性定义。A 型行为模式就是一个间接操作性定义的典型例子。尽管理论概念与其可观测的操作之间联系的紧密程度不同，但所有的概念在一定程度上都需要通过这些与之相联系的实际观测来获得其意义。

科学概念与时俱进

科学概念的定义并非一成不变，而是随着对其观测的不断积累、丰富而逐渐在演变。如果一个概念原始的操作性定义在理论上被证明是无效的，那就要换用另外一组操作来定义它。因此随着相关知识的不断积累，科学概念也在不断的发展并逐步趋向抽象化。例如，有一时期认为电子是绕着原子核旋转的带负电荷的小粒子。但现在，在特定实验条件下，它被看作是具有波浪属性的概率密度函数。

在心理学中，智力这一概念的发展就是一个类似的例子。最开始，这个概念只有一个严格的操作性定义：智力是由智力测验所测到的东西。随着研究数据的不断积累将智力与学生成绩，以及学习、脑损伤及其他行为及生物因素等联系在一起，智力这一概念也在不断的丰富及扩展（Deary，2000，2001；Ferrari & Sternberg，1998；Sternberg，2000；Sternberg & Kaufman，1998；Sternberg & Wagner，1994）。现在看来，最合适的智力概念是由几个更为具体的信息加工操作所定

义的概念所组成的一个抽象构念。这些假设存在的信息加工过程则可通过测量某些行为表现来直接赋予操作性定义。

记忆理论的概念也经历了类似的发展。心理学家现在很少使用诸如“记忆”和“遗忘”这样宽泛的术语，而是去测量几个更为具体、代表不同记忆过程的记忆特征。例如：短时听觉记忆、图像记忆、语义记忆，以及瞬时记忆。最初记忆和遗忘的概念已被细化为具体的操作性概念了。

因此，理论术语的使用是在科学实践活动中，而非咬文嚼字中发展起来的。这是对科学持操作性态度的操作主义及追求纯粹定义的本质主义之间的最显著差别。神经病学家诺曼·戈施文（Norman Geschwind）对这种差异作了如下的描述：“我认为，从医学的发展史中所能了解到的一件事就是：许多人认为研究问题就是要先定义这个问题然后再研究。但这种方法被频频证明是错误的，因为正确定义问题的惟一途径就是先知道答案”（Geschwind，1985，p.15）。

哲学家保罗·丘奇兰德（Paul Churchland）也强调：科学概念的含义是从观察或与其有关的其他概念中获得的，而不是依靠对词语的定义。

要彻底地去理解“电磁场”这个概念，我们必须熟悉其所身处的理论原则体系。只有根据对整个体系的认识才能了解电磁场究竟是什么以及它有什么作用。这个例子是很典型的。理论术语的含义，通常说来，是不能只从单一的、明确的、阐述了所有其被应用的充分必要条件的定义中获得，而往往是由其所身处的理论原则体系推导出其定义（Churchland，1988，p.56）。

科学概念在逐渐的发展过程中，会与不同的理论体系错纵复杂地交织在一起，因而获得不同的操作性定义。这种情形并不代表概念本身有什么错。可惜，许多人不相信心理学，就是因为它的许多重要理论构念，如智力等，拥有许多个不同的操作性定义（Sternberg，2000）。事实上，这种情况并非心理学独有，也不是什么了不起的大事，在科学的领域中这是很常见的。例如，“热”这个术语就可以分别从热力学和动力学这两套理论体系来作不同的构想及定义。物理学也并没有因此而遭到贬斥。再如电子，它的许多特性是从把它视为是一种“波”的角度来加以理解的，而另外的一些特性则又必须从视其为“粒子”的角度来考虑，才能更好地得到理解。到目前为止，还没有一个人会因为物理学存在着这些可相互替代的构想，就提出要抛弃它。

心理学中的操作性定义

提到物理、化学，许多人能够理解操作主义的必要性。科学家在准备研究某一特殊化学反应或是在讨论能量或磁场现象时，人们能够理解科学家们必须用一些方法来测量这些要研究的东西。但遗憾的是，一旦把这些步骤放到心理学的研究上，人们就常常糊涂起来了。要知道，心理学术语也是需要有直接或间接的操作性定义，才能建立一些有价值、并具有解释力的科学理论构念的呀！为什么对于这一点人们总是认识不清呢？

人们对心理学有这个误解的一个原因是，他们对心理学带

有“预设偏见”（在第一章中曾经提到过这个问题）。人们在研究地质学时，不会带着具有情感色彩的信念去探看岩石的性质。然而，他们在看心理学的研究时，情况则大不相同。每个人心中似乎都各自早有一套有关人格和人类行为的理论，因为人们在日常生活中，始终在不断地解释着自己及他人的行为。而且，每一个人的这些个体心理学理论也都含有诸如聪明、好斗、焦虑等理论性的概念，因此人们自然就会问：我已经知道这些概念的意思了，为什么还要让我接受另一些定义呢？尽管表面上看来这种质疑似乎很有道理，但其实这对任何欲理解人类行为的科学发展来说，绝对是个阻碍，同时也是促使公众对心理学迷惑不解的原因。

公众对心理学产生误解，以及媒体在准确介绍心理学发现时所存在的种种障碍，主要是因为心理学的许多专业概念及术语，都是用人们日常生活用语来表述的。这些日常用语让相当广泛的错误观念得以先入为主地进入人们的脑海中。因此，没有受过心理学训练的外行人很少会意识到，当心理学家在使用诸如智力、焦虑、攻击以及依恋等词语作为其理论构念时，它们与自己日常惯用的、叫相同名称的词语的意思是不一定相同的。

在先前对操作主义的讨论中，我们已经可以很明显地看出心理学术语与一般公众所用的术语之间的这种差异的本质。诸如智力或焦虑等专业术语，必须依靠其直接或间接的操作性定义，来决定它们是否在一个心理学理论中被正确地使用。而这些术语常常是极为技术性的及特殊性的，并且在许多方面都不同于这些词语在日常生活中的一般用法。例如，许多人在

听到“对大样本的认知任务进行因素分析所得到的第一主成分因子”这样一句话时，不会意识到这是智力这一心理学术语之操作性定义的一部分。

同样地，“抑郁”这一词语在外行人的词典中，就意味着是“某种陷入低谷的感觉”。而相比之下，《心理障碍诊断和统计手册》(美国心理治疗协会/American Psychiatric Association，1994)，对抑郁病症的几种主要的操作性定义的描述，共占了约12页的篇幅，并且与“某种陷入低谷的感觉”根本就不是一回事。临床心理学家定义的抑郁和外行人认为的抑郁可以说是完全不同的(Hollon，Thase & Markowitz，2002)。其他学科其实也存在同样的问题，只不过不如心理学这么严重。回想一下先前对*生命*这一概念的讨论吧。正如两位米德瓦(Medawar & Medawar，1983)所指出的那样；“麻烦就在于，像科学中许多其他专业术语一样，‘生命’一词是从人们日常生活的用语中借过来的。但是用在科学情境中，其含义已大大不同于其在一般谈话中的用法了”(p.66)。

当心理学家和外行人用同一个词语来指不同的事物时，就很容易会彼此产生误解。如果我们可以创造出一些新名词，来表示心理学家们所研究的构念，那么这种混淆就会少很多。其实，我们时不时也造出了一些这样的新名词。就如物理学家有他们自己的*尔格**和*焦耳*一样，心理学家也有自己的*失调*和*解码*。实际上，这些名词并不是完全是全新创造出来的，只是因为其在日常用语中比较生僻，反而因此能够避免这里所说的混淆。

外行人也许会反对我们发明新鲜名词：“为什么心理学家

* 注：功和能的单位。

要给我们制造这么多麻烦呢？用这么多的行话、生僻用语和高度专业化的定义！我们干嘛要它们？为什么我心中认为的‘智力’就得不到他们的认可呢？”

在这里，我们看到一个对心理学研究有误解的典型例子。这种误解在媒体对心理学研究的报道中经常见到。有一份全国性的报纸以“你能用一般人听得懂的话重新再说一遍吗？”为题，报道了美国心理学联会 1996 年年会的情况（Immen，1996）。这篇文章指出：“心理学家所用的完全是只属于他们自己的语言”。该文嘲讽了在会上报告的一份标题为“用 Gf-Gc 理论来解释结合 WJ-R 及 KAIT 两测验所作的因素分析结果”的论文。虽然这个记者表示“我甚至无法去猜测这个题目到底是什么意思”，但是几乎所有受过严格训练的心理学家都会知道，这个题目指的是这是一篇有关智力测量理论方面之新进展的报告。不过，我们没理由责怪这名记者作出这样的报道，因为我们不应奢望他曾听说过 Gf-Gc 这个名词，因为它是心理学研究智力理论的一个技术上的新进展。就如同我们不会期望该记者知道有关一个基本粒子物理学家刚才发现，但还尚未命名及经过议论的“中微子”的详细资料一样。然而，不知道是怎么搞的，记者对于科学术语的无知（这是完全可以理解的），往往成为人们用以藐视现代心理学的依据。

现在，让我们来看看解决这个问题的关键在哪里。解决这个问题的第一步是要强调在先前的讨论中已提到的一点：操作主义不是心理学研究所独有的，而是所有科学研究都要求有的。大多数情况下，我们很容易接受并认识到这一先决条件是必要的。如果一个科学家发现了一个放射现象，我们理所

当然地认为他（或她）用来测量放射的方法，必须是其他研究者也可以用之来观察到的，并也能以此获得同样研究结果的。操作定义因此导致科学的公开化，而公开化则是科学的关键特征。只要两位科学家就同一个操作性定义达成一致，就可以相互检验对方的结果了。然而，这一个在别处看去很明显的道理，一到了心理学，有时就变得模糊了。人们经常意识不到对如智力、焦虑这样的心理学概念进行操作化定义的必要性。前面说过，原因是我们天天都在使用这些术语，"难道我们还不'知道'它们是什么意思吗？"

这一问题的答案应该是："对！我们确实是不知道它们是什么意思！"这句话中的"我们"当然不是指科学家们，而是指一般普通大众。一个科学家则必须先"清楚"他所研究的智力之含义。并且考虑要如何精确地去定义及测量它，才能让其他实验室的研究者也可以对它进行同样的测量，从而才有可能去看看在针对此概念所做的研究中，是否可以得出相同的结论。这样的操作性定义对具体性和精确性的要求甚高，因此这些科学术语与日常谈话中人们为达到即时的相互理解所用的、意义比较含混的口语用词是大相径庭的。

推动人性化的操作主义

过于依赖所谓"早都知道了，何必再给出新定义"这一心态所引发的问题之根源，同样地也是我们依赖"直觉"（即，非经验）信念体系所会引发的问题之根源。你对某事的"了解"可能与张三、李四所"了解"的有所不同，那我们要如何判断谁是对的呢？你可能会说："可是，我对此

感觉非常强烈，强烈到我知道我是对的。”但是，如果张三跟你的看法有些不同，而他对他的看法的感觉要比你还强烈，那该怎么办呢？然后，如果李四与你和张三俩人的想法又都不同，而他宣称他一定是对的，因为他觉得自己的感觉要比张三的更强烈得多。

这个简单而荒谬的例子只想要说明科学知识的一个基本特性：知识之所以为正确，并不在于它是否是由个别人士对该知识所持有的感觉及相信程度。这个基本特性事实上是推动人类历史实现人性化的主要力量。所有建立在“直觉”基础上的信念体系，都有一个共有的问题：当此类知识出现相互矛盾时，它们没有一种机制来判别哪一个是对的，哪一个是错的。因为人人都是凭直觉感知认为自己是对的，但当大家的感知有了矛盾时，我们该如何判断谁是正确的呢？历史告诉我们，这类矛盾通常，很不幸地，要靠权力争斗来解决。

有些人错误地认为，在心理学中使用操作性手段来做研究是在剔除人性，因此我们应当改用研究者的“直觉”去了解人。心理学家唐纳德·博德本特（Donald Broadbent）提出：所谓研究的人性化，应该在于把关乎人的理论观点，建立在大家都可观测到的行为上，而不应建在理论家的直觉上：

除了观察别人在特定环境中的所说所为之外，我们无从对别人做出判断……实征主义的方法提供了一个摆平分歧的文明途径。如果一个人拒绝走这条道路，那么剩下来的惟一方法就是面红耳赤的争辩了（Broadbent，1973，p.206）。

因此，科学中推动人性化的力量是让知识公开化，以便

用所有参与争议的其他研究者可以用一种大家公认的方法来平息争端。记得在第一章中，我们提到的重验性这个原则吗？这个原则就是允许大家按照预先达成一致的操作机制，在众多相互争议的理论中进行筛选，找出最接近真实世界的一个。而科学要能达到这一公开竞争性，关键要依赖于操作主义这一原则。对概念所下的操作性定义，使得该概念公开化了。也就是说，任何人都可对其进行批判、验证、改进或否定。

心理学概念不能以任何个别人士的定义为基础，因为这种个人定义很可能是很不寻常、古怪或是模糊的。因此，心理学必须拒绝对任何概念作这类个人化的定义（如同物理学拒绝对“能量”作个人定义，气象学拒绝对“云”作个人定义那样），并且必须坚持做可公开的操作性定义，以确保任何一个接受过适当培训并拥有适当设施的人，都可以执行这些操作。拒绝个人定义，不但不会把公众排除在心理学之外，还会把这个学科公诸于众，如同所有学科那样，让大家都能去使用、去探索，并累积出可以为所有人共享的知识。

只有当概念以操作性定义为基础，不理会本质主义者对词义的刨根问底，这种共享的知识才会出现，并被用来解决人类的问题。举例来说，蒙克（Monk）曾描述过二战期间“创伤性休克”这一概念如何成为医学界的一道难题。当时，有些医生确认，创伤性休克的表征为红细胞数量急增，并认为原因在于血液中的血浆漏出，进入了人体组织细胞。另外一些人则认为创伤性休克的表征是低血压、失血多、脉搏快。换句话说，医生对这一概念的操作性定义很不一致（甚至带有个人定义的色彩）。因此一位在英国医学研究学会工作的戈

兰特（Grant）医生建议说：“应当把所谓‘创伤性休克’这个概念丢掉。在不使用这个术语的情况下，对受伤的战士进行详细观察和纪录……由于缺乏大家都同意的诊断依据，致使无法评估不同治疗方法的功效”（Monk，1990，pp.445 ~ 446）。也就是说，这个概念没有一个共享的通用定义，所以使用它不但无益，还可能有害。

科学的概念会随着研究的加深而产生操作性定义的改变。有时，这种改变将会使这一概念在科学中的理解与外行人的理解产生矛盾。费博尔和丘奇兰德（Farber & Churchland，1995）围绕着“火”这一概念论述了这种矛盾的情形。早期火的概念“不仅被用来指涉为燃烧的含碳物质，还包括了太阳及各种星体的活动（实际上是核子的熔解活动），闪电（实际上是由电而产生的炽热），北极光（实际上是辐射光谱），以及萤火虫的闪光（实际上是磷光）。在现代的概念系统中，因为后面这些现象都不包含氧化过程，因此全都不属于木材燃烧的那一类。但是，另有一些现象原本认为是与燃烧没有任何关系的却被证实是具有氧化过程，如生锈、失去光泽以及新陈代谢等，所以可以与木材同属燃烧类。后面这些现象在早期并不认为与燃烧有任何相似之处，是因为当时认为能否感觉到热度是区分是火或不是火的基本特征”（p.1296）。简言之，氧化原则将生锈与营火拉到了一起——而将闪电与之分开，这对科学家来说，可能意味着对火这一概念从事界定的一个新进展。然而，对外行人来说，反而让他们感到迷惑，不知道火到底是一个什么东西了。

本质主义式问题与对心理学的误解

许多人要放弃操作主义这一想法的另外一个原因，是他们之所以想求助于心理学，就是想去探究某些关于人类的问题，希望从中找到本质主义式的答案。这一想法是由于心理学是新近才从哲学中脱离出来的？还是由于，相对于其他学科来说，大众对心理学的了解实在是太少了？究竟是哪一个原因，到目前还说不清楚。就某种意义来讲，是什么原因其实并不重要，因为其最终结果都是一样的：人们似乎期望心理学能够明确地回答出其他学科所不能回答的问题。

回想一下本章开头所谈及的那些问题：地心引力一词的真正含义是什么？其内在本质是什么？在谈到地心引力一词时，我们到底指的是什么意思呢？大多数人会意识到，这些问题关乎现象的本质，亦即其终极性质的问题，现在的物理学理论根本无法解决这类问题。任何一个熟读近几百年来关于物理学发展著作的人都知道，地心引力是一个相当复杂的理论构念。并且，其概念和操作之间的关系也在不断的变化之中。

然而如果把上述问题中的地心引力换成智力，那么突然间，怪事就发生了。上述那些本质式问题的意义一下子就变得重大起来了。这时候问起这些问题也变得那么理直气壮了，而且必须到有关人类终极关怀的本质领域中去探寻答案。因此，每当心理学家像物理学家那样地回答说：智力是一个复杂的概念，它的意义是从测量它的操作以及它与其他构念之间的理论关系去界定的，他们就会被人看不起，并且会被指责是在回避作答。

这是心理学面临的一个难题：公众要求心理学回答那些本质主义式的问题，而对其他学科却通常没有这样的要求。正是这些心理学不能满足的要求，导致了公众对此领域所取得的进展也很轻视。幸亏大部分心理学家（就像其他科学家那样）不去理会及迎合这些要求，而只是专注于自己的工作，因而并未妨碍到这一学科本身的发展。当然，这对公众进一步了解心理学而言，确实增添了不少阻力。所以当一个不了解情况的批评家宣称：心理学没有任何进展时，公众很容易就上当了。这些批评家的言论之所以极少遭到挑战，恰好反映了本书写作的初衷：公众对心理学所取得的实际成就，极度缺乏了解。仔细审视那些对心理学的批评，不难发现它们通常归结于一点：心理学至今没有解决任何本质性的终极问题。对于这种指责，心理学只能说：抱歉，心理学确实没能解决这类问题。但是，别忘了，其他学科也都没能做到。

没有一个学科，包括心理学在内，能够回答本质主义式的问题。很多人可能对这一点感到不舒服。霍尔顿和罗尔（Holton & Roller，1958）曾讨论过，在得知物理学不能回答这类问题时，人们所反映的这种不安。他们所谈的是与放射性衰变现象有关的问题。放射性元素的原子数量的下降与时间是呈指数函数关系的。然而，该函数并不能解释为什么会发生放射性衰变这一现象。要回答这个问题可能必须用到另一个数学函数，但是那个数学函数仍不能回答大众关于放射性衰变究竟是什么的疑问。霍尔顿和罗尔告诉我们：“我们必须接受现代科学的局限性；它并没有声称可以发现所谓‘事情的最终真相’”（pp.219～220）。正如科普作家赖特

（Robert Wright）所解释的：

牛顿的地心引力理论也有些地方不尽人意……然而，究竟“在超距作用”这一构念是否可能在现实生活中找到呢？……牛顿回避了这样的问题……从牛顿开始，物理学家们都以此为榜样……不再试图去解释为什么物体会遵循电磁学或地心引力法则（Wright，1988，p.61）。

同样地，如果那些希望寻求关于人类本性的本质主义问题的人，想求助于心理学，也必定要失望的。心理学不是宗教，它是追求用科学的方法来理解行为各方面的一门宽广的学科。因此，心理学目前对行为所提出来的解释还只是一些暂时比起其他理论来说，是更贴切的理论构想。将来这些理论构想肯定会被更接近真理的构念所替代。

操作主义与心理学对问题的措辞

操作性定义的方法对评估心理学理论的可证伪性，是一个非常有用的工具。概念是否以直接或间接的可观测操作为基础，是判别理论有无可证伪性的重要线索。那些没有这种基础的概念通常是为了配合收集后的数据，而刻意构造出来的，为的是要避免该理论遭到否定。因此，对待那些理论家不能给出直接或间接操作性联系的不严谨概念，我们应抱有怀疑态度。

这与科学家们称之为“节约”（parsimony）法则有关。节约法则认为，当两套理论拥有同等解释力时，更简单的那套理论（含有较少概念和概念之间关联的那套）更可取。原因是，含有较少概念性关联的理论，在以后的测试中会更具

有可证伪性。

对操作主义原则的深入理解有助于我们找到或辨别出哪些是没有什么科学意义的问题。例如，在我的文件夹中有一篇来自合众国际社（United Press International），名为“动物可以思考吗？”的文章。它报道了最近对动物行为所进行的一些实验。文章所述的研究没有任何不妥，但很显然的，这个题目却是在耍花腔。除非可以指明“思考”这一术语的操作性标准，否则题目中所问的问题就毫无科学意义的。但是，在该文中并没有提到操作性定义这个问题。许多最近的报纸刊文也存在类似的情况：“计算机能思考吗？”。同样地，没有对思考定出操作性标准，这类问题就无科学意义可言。不过，这些问题在鸡尾酒会上，倒是可以大派用场。

事实上，观察人们在最后一个问题上的争论很有些教育意义，因为这种争论使我们有机会具体见证，先前讨论的那种、对心理学所持有的 “预设偏见”问题。大多数人都有一个强烈的预设偏见，不想让计算机也能思考。为什么呢？主要原因在于：大众对“思考”的想法，是与“人类”这一概念，紧密纠缠在一起的，以至于许多人在情感上不能接受“非人”的东西也能思考（例如计算机，或看起来不像我们地球人的外星生命体）。

然而，尽管人们强烈抵触计算机能思考的想法，但是这并不意味着大多数人对这个问题曾经仔细考虑清楚。对“思考”的定义他们常常是稀里糊涂地把绝大多数人（如婴孩）都拉了进去，而将所有计算机都排除在外。没有跟进目前人工智能进展的人，对“思考”所提出的定义标准，有时听

来颇为滑稽，因为他们总是恰恰选择了那些计算机都能胜任的事情作为他们的标准。例如，许多人提出思考是“从经验中学习的能力”，那么我们只能告诉他们，用计算机运行的许多人工智能系统都符合这一标准（Churchland，1995；Clark，2001；Elman，Bates，Johnson，Karmiloff-Smith，Parisi & Plunkett，1996；Pfeifer & Scheier，1999）。然而，千万不要小看“预设偏见”的力量呀！你可能以为人们接下来会这样反应：“噢，我不知道计算机能做这些事。既然计算机符合我提出的‘思考’标准，那么我不得不说至少有些计算机是能够思考的吧！”哈！你错了，通常大家是不会做出这样理智而诚实的反应的。更常见的反应是煞费苦心地去重新另找一些标准，希望计算机这次不能达标。

通常，对“思考”所下的第二个定义标准是 “创造性”（“想出人们认为有用，但先前没有其他人想到的东西”——让我们暂不考虑去回答大多数人是否能够达到这一标准，这一问题）。但即便大多数专家同意计算机符合这一标准（Boden，1990；Pfeifer & Scheier，1999），仍有人不肯承认机器有能力思考的可能性。这时，他们往往不再求助另一个操作性定义，而是直截了当地争辩：计算机不可能会思考，因为“是人发明了计算机，并且为它编写程序；计算机只不过会执行程序而已”。

尽管这种言辞早已成为攻击机器可以思考的陈词滥调（McCorduck，1979；Robinson，1992；Woolley，2000），但它仍具有骗人的本事。“预设偏见”使许多人意识不到这些辩解与要讨论的问题毫不相关。几乎人人都会同意，思考

是自然界产生的一个过程。请注意我们在讨论其他自然界的过程时，都没有涉及“起源”的问题。试想一下食物加热的过程，并问这样一个问题：“是炉子在加热吗？”我们会不会说：“真正在加热的不是炉子，因为它是人造出来的。因此只有说是人在加热食物才有道理”？那么轮到起重机时，我们又该怎么讲呢？“是起重机在举重吗？”我们的回答是不是还是“不是起重机在举重，因为它也是人造出来的，所以我们只能说是人在举重”呢？当然不能这样说。事物的起源与其执行什么特定步骤的能力无关。思考的过程也是如此。一个实体能否思考并不依赖于该实体是谁发明制造的。

针对人们无法理性地对待机器有可能会思考这一问题，计算机专家艾伦·图灵（Alan Turing）设计了一个测试计算机能否思考的著名实验。这一实验对于我们当前的讨论尤为重要，主要是因为图灵所设计的测试属于操作性测试。图灵以“我提议考虑这样一个问题‘机器会思考吗？’”为开场白，开始了他那篇著名的论文——《计算机和人工智能》（1950）。图灵对计算机能否思考的探研，不是鸡尾酒会上的那种无聊闲扯，也不是本质主义对“思考”的词义作无休止的争论，而是提出了一个严格的操作性测试：如果计算机能够进行理智对话的话，那就可以说它是有思考能力的。

图灵这一提议的创造性在于它既提出了操作这个思考概念的方法，同时又防止了“预设偏见”的干扰。他严格地明确了测试的每一个环节，该测验并不是简单地把一个计算机终端摆在屋里，让一个被试用鼠标和键盘与它对话，然后判断计算机有没有和这个人进行理智对话。图灵的测试不是这样简

单，因为他很担心这一测试过程存在“预设偏见”的干扰。他确信，如果让一个人坐在屏幕和键盘前——这个人很显然地会意识到这些设备都是机器，那么不管计算机是否真有思考的能力，他（或她）都会否认机器会有这个能力。因此他建议，测试应控制那些与思考能力无关的外在因素。他的著名实验设计是，让被试同时与两个键盘对话，其中一个键盘的一端是一部计算机，而另一个键盘的一端则是一个活生生的人。当然被试并不知道这两个键盘后面哪一个是人、哪一个是计算机在操作。被试的任务是在通过键盘与之对话后，去判断哪一个键盘后是计算机，哪一个是人。如果被试不能以大于机会概率的正确率猜出哪一个是人，那么我们就有理由推断计算机具有同人一样的对话能力——这正是“思考”的操作性定义。

图灵的设计“与交响乐团选拔乐师的测听面试之设置有异曲同工之妙。试听面试时，在评委与面试乐师间要放一个不透明的屏幕，前者是要隔着这个屏幕来判断乐师演奏的好坏。显然地，挑选演奏家，我们关心的应该只是他（或她）的音乐能力，如此而已。诸如性别，头发长短，皮肤颜色，体重等这些外在因素应该是毫不重要的，所以必须将它们摒除在评判之外才好……图灵认识到人们对智力的判断，同样会因对方是不是有柔软的皮肤、一腔热血、有面孔、手、眼，而产生偏见，而这些特征显然都不是我们所构想的智力之基本成分”（Dennett，1998，p.5）。图灵的测试告诉我们：如果我们要理性地讨论心理学概念，而不只是反映我们自己对某个问题的偏见，那么操作性定义就是必要的。

人们在讨论人工智能问题时，所使用的思维方式，很好地

说明了科学与非科学思考方法之间的差异。科学的方法是先发展出一个合理的操作性定义，然后再看，不管是从“计算机”或是“人”那里，我们可以得到哪些有关“思考”的结论。然而，预设偏见却左右了一般大众的思维方式。这些人早已对这一问题有了先入为主的结论，所以对“机器”与“人”的真正差别到底在哪里毫不感兴趣。一旦形成定势，他们就会绞尽脑汁拼命地去找出各种理由，来稳固自己的这些想法，不让它们改变。但是我们已经看到，正是这些预设偏见，加上非操作性的本质主义的心态，滋长了人们认为无须任何必要的操作性标准，就能“知道”思考“真正”是怎么一回事了。也正是这种心态使得大多数人的“直觉”心理学理论无从被证伪，因此是毫无价值的。这种心态也清晰地向我们展示，为什么我们需要为一门科学的心理学！

小　　结

操作性定义是指，以可观察、可测量的操作给一个概念下出来的定义。确保一个理论具有可证伪性的主要方法之一，就是要弄清楚理论中的关键概念是否具有操作性定义，以及该操作性定义是否具有良好的可重验性。操作性定义是科学知识得以公开并接受检验的一个主要机制。通过这样定义下来的理论概念被公布于众，可以任由所有人去重验——不像直觉性定义是一种非经验性的定义，只属于个别个体所特有，因此无法公开地让其他人去验证。

由于心理学所使用的术语多源于日常话语，像智力、焦虑

等，因而许多人对这些术语的含义都会有预先惯用的想法，所以对这些术语进行操作性定义的必要性，常常得不到人们的认同，认为是多此一举。其实，心理学跟其他科学一样，也是需要对其术语进行操作性定义的。然而，人们往往不这么想，他们经常要求心理学去回答一些本质主义式的问题（关于概念的绝对、根本实质的问题），却对其他学科不作如此的要求。事实上，没有任何一门科学能够回答这类一成不变的终极性问题。心理学与其他科学一样，其操作性定义也是在不断地修改之中，以致使其理论概念能更加精确地反映世界的本貌。

（潘伟 译）

第四章　见证和个案证据

——安慰剂效应和神奇兰迪

在这里，让我们将电视频道切换到近十年来美国最受欢迎的一个脱口秀节目“温佛瑞秀”（*Oprah Winfrey Show*）。今天的嘉宾是阿尔弗雷德·庞提菲科特（Alfred Pontificate）博士。他是俄狄浦斯人类潜能研究所所长。主持人走到观众中间，要大家就博士刚刚提出的、目前尚存在不少争议的“出生顺序理论”进行提问，这个理论的基本思想是：一个人的生活历程早已在其与家人的互动中被定型，而与家人互动的型式又是由其出生顺序所决定的。无一例外地，像这种讨论的话题必然会从理论的探讨，转向对个别观众之重要个人事件的解释上。对此，这位博士也欣然遵命了。

例如，有人问博士：“我哥哥是个自我毁灭式的工作狂，他不关心妻子，也不在乎家庭，只考虑与自己工作相关的事情。他患有溃疡，还有酗酒问题，但他拒绝承认。他和他的家人已经两年没像样的一起度过假了。他逐渐在朝离婚的悲剧走去，但是他好像根本不在乎他的婚姻。究竟为什么他会选择这样一条自我毁灭的道路呢？”

对这个问题，博士回应：“请问，他排行第几？”

“哦，他是老大。”

“这就对了，”博士说，“这是很常见的。临床中我总

是见到这样的个案。这类情况产生的原因是，父母把他们的希望和挫折都转移给了他们的头生子。这是一个无意识的过程，虽然父母没有明说，但孩子还是会将这些希望和挫折内化成为对自己的评价标准。我把这一无意识过程叫做*期望上旋*的动力过程，通过这一过程，父母的愿望就会转换成为孩子对自己追求成就的过分、甚至病态的需要。”

尽管有时候，当嘉宾的观点与观众的信念相左时，观众会提出一些尖锐的问题；但是如果一个所谓的行为“专家”是一味在附和观众的传统惯性思维，这种针锋相对的情况就不太会发生。不过，偶尔也还是会出现这类火爆的小插曲，这时场上的气氛就热闹起来了。今天，现场就有这么一位热情、直率的提问者：“请等一下，博士，我哥哥也是头生子，我父母把这个游手好闲的家伙送到哈佛去读书，而只把我送去读一个两年制的专科学校学牙科。一年后，我父母心中的这位神童退学了，去了科罗拉多的山区。我们最近一次见到他时，他正在那里编篮子！你所说的头生子现象，我很不明白，好像与我哥哥的情况完全不同。”

观众席的气氛顿时因为这一挑战而显得紧张起来，但是，不打紧，这位博士总能给观众一个令人满意的解释的：“哦，是的，我见过很多像你哥哥这样的个案。对，我在治疗的过程中，经常会遇到这样的一些人，他们的*期望上旋*的动力过程已经中断，因而产生了一种无意识的愿望去阻止将父母的期望转移到自己身上。这样，个体的生活规划就会朝着排斥传统的成就标准的方向发展。”一阵肃然沉默之后，下一个“个案”又开始登场了。

我们以上所述说的，当然都是大家相当熟悉的场景。这又是一个“拉什问题”的例子。这个出生顺序“理论”是在没有事例可以证明“它是错的”的思维框架下被构想出来的。它是一个不可能被证伪的理论，对它进行实征研究是没有意义的，因为没有任何事情是不可以用这个理论来解释的，或是排除于它的理论预测之外的。

然而，本章所关注的问题并不是这个理论本身，而是用以支持这个理论成立的证据之性质及质量。庞提菲科特博士用他自己的“临床经验”或“个案研究”作为支持其理论的证据。这在传媒心理学是经常可以见到的。在脱口秀节目里及书店一般普及读物的书架上，到处都是以创始人的临床经验为基础而成立的心理学理论。通过这些媒体管道，一些展现在大众面前的治疗法都仅仅是出自接受过该项治疗，并认为这种治疗有效的个体的亲身口述。这种口述通常被称为是“见证”(testimonials)。在这一章里，我们主要想说明一个道理，这个道理可以给想知道要如何善用心理学知识的人提个醒：用个案研究结果或见证叙述作为评价心理学理论和治疗效果的依据，几乎可以说是完全没有价值的。

下面将阐明为什么这些证据不适合用来支持理论及疗效，并讨论个案研究在心理学研究中的作用到底在哪里。

个案研究的价值

个案研究信息的价值，在很大程度上取决于某一领域科学研究的进展。在对一些问题研究的早期阶段，从个案研究或临床

经验中所获得灵感与启发是非常有用的，它们提示我们哪些变量值得进一步研究。个案研究对心理学开辟新的研究领域曾经起过非常关键的作用。皮亚杰（Jean Piaget）的工作就是人尽皆知的一个例子。他通过研究，提出了儿童的思维不是成人思维的简易版或初级版，而本身是自成一体的设想。皮亚杰关于儿童思维的这一设想，有些已经得到证实，但是很多到目前还没有被证实（Bjorklund，1995；Goswami，1998）。不过，在这里讨论的重点不是他有多少设想得到证实，而是要理解：皮亚杰的个案研究没有证实任何东西，它只是为发展心理学提供非常多的、有待进一步研究的亮点。真正为皮亚杰从个案研究中提出的假设，提供了证实或证伪之证据的是随后所进行的一些相关研究和实验研究（将在第五章和第六章讲到这些研究）。

然而，当我们从科学研究的早期阶段过渡到比较成熟的理论检验阶段时，情况就彻底改变了。个案研究在科学研究的后期是发挥不了什么作用的，因为这时主要的研究工作是在检验特定的理论，而个案研究不能提供能让这些理论被证实或被证伪的证据。其原因是：个案研究和见证叙述都是所谓的孤立事件，缺乏必要的比较信息，来帮助排除对同一事件或例证的其他可能的合理解释。

见证叙述之所以和前面给的个案研究的例子一样是孤立事件，因此是不能验证对错的，是因为见证叙述通常只包括支持某一理论、治疗或事件的话语。这样一来，问题就来了，如果每一个试验疗法都可以累积一些支持其疗效的话语，那么，我们又如何可能用之判定任何一个疗法是好或坏呢？正如心理学家雷·尼克尔松（Ray Nickerson）在评论我们常常用以自欺

的认知过程时，曾说过："江湖郎中的骗术往往得逞，是因为他们随时可以找到一些病人愿意为他们做见证，这些病人总是满怀诚心地，告诉他人自己的确在治疗中有所康复。"（Nickerson，1998，p.192）。例如，一些卖潜意识自助录音带（那些用一种低于听觉阈限的信号制作出来的录音带），往往以大量的用户给出见证证据，声称可以提高人的记忆力、甚至自尊。然而，在严格控制条件下所做的研究显示：这类录音带对记忆力或自尊没有任何改进*（Greenwald，Spangenberg，Pratkanis & Eskenazi，1991；Moore，1995）。

理解理论检验，最关键的是要认识"其他可能的合理解释"这一概念：一个事件或现象通常可以用许多相互可替代的理论来解释**。实验设计的目标就是组织安排这一事件或现象的出现，使其只能用某一特定理论才能解释，而用其他理论则解释不通。正如第二章所说的，只有利用我们精心安排的实验程序，排除了其他合理解释的可能性，我们所收的实征数据才是有价值的，科学研究才会有进步。科学可以说是为各种相互竞争的理论或观点提供了一个自然选择及淘汰的途径。有些解释经过实征检验存活了下来，而另一些则被淘汰出局。保留下来的可以说都是比较更接近真理的。但是，这是个慢工出细活儿的工作，各种理论观点必须要反复接受审查，以便发现哪一个最近真理。同时，这必须是一个"你

* 译者注：目前在国内的电视上，最常见的类似情况就是一些出售药物或理疗仪器的公司，找来用户在广告里大肆宣扬自己如何在用其他方法百治无效的情况下，能迅速得以治愈的见证叙述。

** 译者注：例如一个人为什么会自杀，可能是因为他的自我评价太差，或是之前被父亲打了一个耳光。这两个解释因为都有可能，故称为是相互可替代的理论。

死我活”的竞争过程：为支持某一特定理论所收集的证据，不能同时也可以用以支持其他同等可能的合理解释。基于这一原因，科学家在他们的实验中设有控制组，或称为比较组，以求得到比较信息。这样做的目的，是在对控制组的结果与实验组进行比较时，其他可能的合理解释可以被排除。实验设计如何可以做到这一点，是后面几章要讲的内容。

个案研究和见证叙述因为没有给出比较信息，只代表一些偶发的孤立现象。它们缺乏必要的比较信息来证明某一特定理论或治疗是更具优势的。因此，引用见证叙述或个案研究来支持该特定理论或治疗，是一个错误的举措。这样做的人，如果不向公众挑明他们所提供的这类所谓证据，很可能可以用许多其他的合理解释也解释得通，那么他们就是在误导群众了。总之，对现象的这种单一孤立的论证具有极高的误导性。下面将用安慰剂效应来更具体地说明这一点。

为什么见证叙述是没有价值的：安慰剂效应

实际上，医学和心理学所实施的任何一个治疗都能获得一些支持者，总会有一些人诚心诚意地来证实其效果。在医药科学的记载中，有人曾见证过，猪牙、鳄鱼粪便、埃及木乃伊制成的粉、以及其他更奇奇怪怪、富有想像力的“药物”都具有疗效（Harrington，1997；Shapiro，1960）。但实际上，我们早就知道：仅靠向病人暗示某一治疗正在进行之中，就足以让许多病人感到病情好转。

不管治疗是不是真的有效，人们通常都倾向于报告它们有

效，这就是著名的安慰剂效应（Christensen，2001；Dodes，1997；Ernst & Abbot 1999；Harrington，1997；Kirsch，1999；Russo，2002）。在电影《绿野仙踪》中，这个概念得到很好的展示。巫师实际上没有给锡人一个心，给稻草人一个脑，给狮子以勇气，但是他们都感到比以前要更有爱心、更聪明，或更具勇气。实际上，由于直到近一百多年，医药科学才发展出较多的、具有确凿疗效证据的治疗来，有人曾经这样说："在这个世纪以前，医学史只能说是安慰剂效应的历史罢了"（Postman，1988，p.96）。

我们可以通过考察生物医学研究来说明安慰剂这一概念。在生物医学研究中，所有的新药研究程序必须包括对安慰剂效应的控制。一般来说，如果在一组病人身上试验一种新药，就要组建一个患同样病症的对等组，给他们服用不含该药的等量药剂（安慰剂）。两组病人都不知道他们吃的是什么药。这样，两组进行比较时，安慰剂效应——即给予病人任何一种新的治疗都会使他们感觉好些的这种倾向就可以得到控制了。仅仅报告百分之几的病人吃了新药后症状得以缓解是不够的，因为如果没有控制组的缓解百分比资料，就不知道这百分之几的病人，报告其症状缓解是由于安慰剂效应还是药物本身的效果。对各种疾病患者的研究（见 Bok，1974）显示，大约有 35% 的人，吃了安慰剂后，病情也获得令人满意的缓解（也见 Roberts et al.，1993；Russo，2002）。最近的一个研究认为：与常用抗抑郁药氟西汀（Prozac）相对应的安慰剂效应，是该药本身效果的两倍（Bower，1996b）。甚至有发现，安慰剂效应可以强到令病人对安慰剂药片上瘾的地

步（Bok，1974；Ernst & Abbot，1999），从而必须依靠吃安慰剂来维持他们的健康！

当然，在对药物疗效所做的对比研究中，对安慰效应的控制事实上并非给病人一个完全没有任何疗效的药片。相反地，它是当前认为能控制该病的最好的药片。实验比较所要揭示的问题是，新药是不是比当前最有疗效的药还要好。

安慰剂效应在心理治疗中也很常见（Wampold et al.，1997）。许多有轻度和中度心理问题的人，在接受心理治疗后，说他们的情况有所好转。然而控制研究证明：这一康复比例中有相当一部分人，是由于安慰剂效应和单纯的时间推移这两个因素共同作用的结果，时间推移通常被称为是自然康复现象。大多数有疗效的治疗都是治疗效果和安慰剂效应的结合。正如多德兹（Dodes，1997）所说，“即便非常严重的疾病也有恶化与康复交替出现的过程；关节炎和多发性肌肉硬化是最常见的例子。甚至连有些人患有肿瘤，也会莫名其妙地消失了”。不过他也提醒大家：有安慰剂效应，并不意味着病人的问题是无中生有的。和一般看法认为安慰剂是无害的正相反，他警告大家说：“安慰剂效应往往会证实和（或）强化人们怀疑自己有某一疾病的想法，从而‘诱发’出他们的慢性疾病”。从而使他们变成毫无科学根据、只靠安慰剂效应，来从事治疗的江湖郎中之常客”（Dodes，1997，p.45）。

在心理治疗疗效的研究中，通常很难决定到底要如何来对待服用安慰剂的控制组，才能让他们给我们所需要的比较信息。不过，让我们先暂且不谈这个复杂性问题，而是去理解

一个更基本及重要的问题：研究人员为什么要把治疗效果与安慰剂效应和自然康复现象区分开来。保罗（Paul，1966，1967）的一项对心理治疗疗效的研究，提供一个很好例子，让我们看到这类研究能够告诉我们什么。保罗研究了几组在公共场合说话会感到不安及焦虑的学生。其中实验组接受了专门针对说话焦虑问题的抗敏治疗，有85%的人得到明显改善。安慰剂组吃的药被告知是一种强力镇静剂，但实际上只是小苏打胶囊。这一组中，有50%的人说有明显改善。第三组没有接受任何治疗，也有22%的人说他们的问题得到明显改善。从这一研究中，我们可以看到，这一心理问题的自然康复率为22%，由于安慰剂效应而得到改善的百分比为28%（是50%减去22%），而抗敏治疗的疗效则为在安慰剂效应之上的35%（85%减50%）。

同保罗的研究结果一样，其他研究也表明：心理治疗的疗效确实要好于只用安慰剂所可能产生的效果（Lipsey & Wilson，1993；Nathan & Gorman，1998；Prioleau，Murdock，& Brody，1983；Smith，Glass，& Miller，1980）。但是使用安慰剂为控制组的实验也表明：仅仅用实验组与控制组病人自我报告感觉有改善的人数之百分比的差异，会严重高估该治疗的实际疗效。问题在于见证叙述实在是太容易得到了。康奈尔大学心理学家吉勒韦克（Gilovich，1991）曾指出："由于人体自身有很好的自我康复能力，即使医生什么管用的事都没做，病人中有很多也会感到病情有所好转。因此，当因自然康复令治疗率的基线（即上述的50%线）变得很高时，甚至一个完全没用的治疗也会显得很有效"

（p.128）。总之，只要我们进行医疗干预，安慰剂效应就无时无刻不在，不管这种干预本身是否真的有效。由于安慰剂效应是如此之强烈，哪怕治疗方法再荒唐滑稽，只要它是向一大群人实施的，总有几个人愿意出来见证它的疗效（“清早击头治疗法——每天早起敲打你的头部，你会一天精神百倍！邮寄10.95美元给我，你就可以得到这个特制的、经过医学测试过的橡胶槌”）。

对不起，针对这么一个严肃事情，我们真的不应该开玩笑。毫无理由地去依赖见证叙述和个案研究证据很可能带来灾难性的后果。例如，曾为抽动秽语综合征做出科学构念化——将之界定为器官障碍的研究小组（Shapiro et al.，1978；见第二章），其成员曾经指出，由于人们不恰当地依赖个案研究证据，使得对该病的病因一直停留在不可被证伪的心理分析理论框架内打转，阻碍了对该病病理进行真正科学研究的进展。英国科学杂志《自然》驻华盛顿记者巴蒂安斯基（Budiansky，1984）在总结医学领域中的这种现象时，曾说了下面这段话，这段话更强化了前面几章曾经讨论过的，关于什么是科学的许多要点：

科学回避所有个性的东西。虽然这一偏向常常被归因是出于科学家们应有的冷静，但事实上，它确实是20世纪人类智慧的伟大胜利品之一。科学家们深知，要寻找自然界的原因和结果，就必须将个人的感觉及经验排除。健康科学曾为此吃尽了苦头；人们生病有许多原因，大多数情况下，人们无论在接受什么“治疗”后，都会有所好转。所以，看来永无止境的各类庸医庸术，凭借着那些对疗效满意患者的见证叙

述，总是广受人们爱戴。这些就是活生生的例子，说明要我们跳出个人经验的束缚有多么困难（p.7）。

《新英格兰医学杂志》的一篇社论，反映了医学界对个案研究及见证叙述在医学研究中可能发挥的作用："比方说，如果我们杂志接受了一篇论文，是叙述一个患胰腺癌的病人在服用了大黄根（rhubarb）后得到了康复……我们可能会将之作为一个案报告来发表，但是我们之发表它并不是宣告它为一种新的疗法，而仅仅是建议它为一个值得用正规的临床实验测试去进行验证的假设。与此正相反，那些讲述各类秘方之疗效的文章（通常发表在通俗的书籍和杂志上），则没有做出这样的声明，因此这些文章本身也就不足以作为对那些疗效的支持文献了。"（Angell & Kassirer，1998，pp. 839 ~ 840）。

"鲜活性"问题

前面已经阐明因为有安慰剂效应，致使见证叙述无法成为有用的科学证据。能指出这一点固然是好，但是这还不够，要人们摒弃见证式证据还须扫除另外一个障碍。让我们先对这一障碍有所认识。社会和认知心理学家都曾对人类记忆和决策过程中所出现的鲜活性效应（vividness effect）作了许多研究（见 Baron，1998；Kunda，1999；Ruscio，2000）。当面临解决问题或作决策的情境时，人们会从记忆中提取与当前的情境有关的信息。因此，对解决当前问题而言，越是容易

获取（亦即具有可获性）的事情，越会被人们用之来解决问题或作出决策。而信息或事情的鲜活性就是强烈影响信息之可获得性的一个因素。

这样，问题就来了，因为没什么会比人们真真实实地叙述他们看到的东西或经历过的事情，更加具有鲜活性、更加能打动人心的了。个人见证的鲜活性经常使得其他更可靠的信息变得黯然失色。有多少次，我们在购物前虽然仔细收集了大量不同产品的信息，却在最后时刻，仅因一个朋友或一个广告而改变自己最初的购买决定？购买汽车是一个典型的例子。大概在翻看了《消费者报告》中的数千份消费者调查之后，我们终于决定要购买一辆X车。又参考了几本汽车杂志，里面的专家们也推荐X车，这更坚定了我们的选择——直到，在一次聚会上我们遇到一个朋友，他说他朋友的一个朋友买了一辆X车，结果是一辆烂车，老有毛病，花了几百美元去维修，而且决定再也不会买这个牌子的车了。单单这样一个事例本不应该那么严重地影响我们对X车的观点，因为我们是在收集了针对数千名用户所做的调查报告和众位专家的评判之后，才决定要买X车的。然而，我们中究竟有多少人会不受这一见证叙述的影响呢？

以上这一买车的事例说明了：鲜活见证证据所产生的问题不仅只存在于心理学这一领域。鲜活性影响人们观点的事例处处可见。记者约翰逊（Johnson，1991）曾记载美国里根总统是如何认识到艾滋病之严重性的事例：

里根直到1985年才认识到艾滋病的严重性，当时他看到一则新闻报道说，演员洛克·赫德森（Rock Hudson）死于

这一病症*。而这时艾滋病已出现了5年之久，数以千计的美国人遭到感染，关于它的话题也早已传遍了街头巷尾。但里根是看到赫德森的死亡报道时，才要赫顿将军（以前作过医生）给他讲讲这种疾病（p.454）。

简言之，新闻报道和统计数据都显示艾滋病死亡人数在不断增加，但是这些资料并没有能够吸引总统的注意，但他所认识的朋友中，仅仅一个人的死于此症就已经引起了他对该症的注意了。同样，作家迈克·刘易斯（Michael Lewis）也讲述了著名保守派时事评论员乔治·威尔（George Will）——一个坚决反对强势政府干预老百姓生活的人——是在他家门外看到撞车造成死亡事件后，才发表专栏文章呼吁政府要强制所有汽车安装防碰气囊（Lewis，1997）。

试想一下，在一个周五的早晨你在报纸上看到一个大字标题："巨型客机坠毁：413人丧生"。你可能会想:天啊，多可怕的事呀！怎么会发生这样悲惨的事情！再试想，过了几天之后到周四，你一起床，看到报纸上说："另一架巨型飞机遇空难：442人死亡"。那时你可能会想:唉呀!不可能吧!多可怕呀!我们的空中交通系统到底出了什么问题？接着再试想，又到下一个周五，报纸上又出现："第三个撞机悲剧：431人命丧黄泉"。这时，恐怕不仅是你，全国都会震动起来，联邦政府会被要求尽快调查此事，飞机会被停飞，许多调查委员会将设立，并会引来一大堆的诉讼案件。《新闻周刊》和《时代》杂志会以此为封面报道，连续几天这件事都会是

* 赫德森是著名的好莱坞演员，里根在当演员时和他是好友。

电视新闻节目的头条，电视记录片也会对此作详尽的分析。一时间，硝烟万丈。

请注意，这里叙述的现象并不是虚构的，它所带来的问题都是真的。每一周的确有一架巨型飞机坠毁。好吧！也许不是一架巨型飞机，而是许多架小飞机行不行？。不行，好，不是小型飞机，而是小型交通工具可不可以？这种小型交通工具叫机动车。全美国每周大约有843人死于交通事故（根据国家安全委员会2001年的报告，每年共43,500人）。死者足以坐满一架巨型飞机，这总可以了吧！

每周死在高速公路上的人，与一架巨型喷气机坠毁所死亡的人数相当，但却没有引起我们的注意！这是因为媒体没有将其以“正如一架满载客人的巨型飞机坠毁”这类鲜活的方式呈现给我们的原故。所以，每周死于小汽车交通事故中的457人（加上另外死于卡车和摩托车事故的263人），对我们来说，没有鲜活性。我们不会在餐桌上像谈论飞机坠毁、众人死亡那样地谈论它们。不像如果飞机真的每周坠毁、每次死亡400人，我们一定会讨论空中交通系统的安全性那样，我们也没有去争辩汽车行驶的安全性和必要性。媒体没有鲜活地给我们呈现这457人的死亡，是因为这些交通事故不是发生在同一时间、同一地点，因此媒体只能（偶尔地）给我们提供一些数字（如，每周457人）而已。这一数据本应足以引发我们思考，但是却没有。根据调查数据，驾车是一项极端危险的活动（Lichtenstein，Slovic，Fischhoff，Layman，& Combs，1978；国家安全委员会，1990，2001）。然而，至今从未见有就这一活动所带来的危险性，相对于其所带来的

便利，进行任何全国性的大辩论。难道这里所说的死亡数字，就是人们为了享有当今的郊区生活，因而需要大量驾车往返于工作及居住地点之间，所必须付出的昂贵代价吗？我们从来没有提出过这类的问题，因为我们从来没有认识到这是一个问题。因为它所付出的代价没有像空难那样，是以鲜活的方式呈现给大家的。

请大家再试想下面的一个例子，看它有多么滑稽。你的朋友开车20公里送你去机场，你从那里将乘飞机到一个750公里以外的地方。离别时，朋友可能对你说："一路平安"。这句话可以说是既伤感又颇具讽刺意味。因为他（或她）在回家的20公里车程中，死于交通事故的几率与你所飞的750公里航程的坠机事故几率相比，高出3倍多（据国家安全委员会1990的报告）。"A祝B旅途平安"事实上是一个非常不理智的祝福，因为A的处境事实上比B更加危险！然而，受到鲜活性效应的影响，让我们觉得A给B的祝福是合理的（Sivak & Flannagan，2003）。

科尔（Cole，1998）在一次环球航空公司空难后所做的民意调查中，发现如果飞行安全系数能够提高，绝大多数被调查者愿意为一张双程票多付50美元。然而，同样的这些接受访问的人，却拒绝支付比50元少得多的钱去购买一些汽车内的安全措施，这些措施可以带给驾驶者更高的安全保障。同样地，很少人知道：他们在驾车10公里去买一张彩票时，其死于交通事故的几率比其中奖的几率要高出16倍（Orkin，2000）！

依据大众媒体所展示的鲜活形象，误导人的个人判断可谓

随处可见。一项研究报告（MacDonald，1990；也见Cole，1998），针对父母作了调查，要他们举出最担心孩子遭遇什么危险。结果显示父母最担心的是孩子遭绑架，而这一事件的几率是1/700,000。与之相对照，孩子在车祸中身亡的危险，父母则不太担心。然而这种可能性比遭绑架的可能性要高出100倍（Paulos，1988）。显然，遭绑架的担心大部分是媒体渲染的结果。这项研究结果说明，由于鲜活性效应，带来感知上的扭曲，“美国父母所担心的竟是一些不大可能发生的事”（MacDonald，1990）。一个研究人员哀叹道：将担心聚焦在“当下流行”的事件上，必然会误导父母的注意力，以致使他们“忽视了更有可能造成坏影响的事件，例如学业成绩、看电视的坏习惯、吸毒和驾车安全等”（MacDonald，1990）。

作家彼得·博伊尔（Peter Boyer）曾描述维护带枪自由的说客们如何以同样的方式助长美国人对危险的错误认识（Boyer，1999）。这些人试图让公众的注意力集中在非法入侵者破门而入的鲜活案例上。他们的潜台词是：这时，如果有一支枪来保护自己，你就没危险了。但博伊尔（1999）指出，极具讽刺意味的是，枪支工业的意图是去把焦点放在“坏人手中的枪”上，而实际的统计数字显示：在现实生活中，问题往往出现在“好人手中的枪”上。这是因为犯罪并不是这个国家大多数的枪击死亡案的主要原因，事实上死于用枪自杀的人要比死于枪杀案的人多得多。多数枪击死亡可归因于枪支走火和自杀。这正是为什么这一研究的最终结论是：把枪带回家反而增加了家庭的危险（Miller，Azrael &

Hemenway, 2002)。

单一个案的压倒性影响

虽然人们有许多正确信息可以利用来帮助他们作判断，但是他们往往不顾这些信息，单单要被一个单一的、突出的例子所左右。对人们这一倾向，心理学家曾展开了深入的研究。威尔逊和布瑞奇（Wilson & Brekke, 1994）证实了鲜活性的危害，以及它对消费者行为的实际影响。他们调查了两个不同品牌避孕套（品牌A和品牌B）的两类信息对人们产生的不同影响。一类信息是《消费者报告》杂志上的一次调查及分析。另一类信息是由两名大学生叙述他们对某一品牌的偏爱。首先，威尔逊和布瑞奇调查了一组被试：询问他们更愿意接受哪一类信息的影响。超过85%的被试说，他们会选择《消费者报告》所给出的信息，而不是那两名大学生所给出的观点。然后，威尔逊和布瑞奇选择了另外一组与前一组相匹配的被试进行实验，他们告诉被试可以免费得到自己选择的避孕套，而且被试在选择的过程中，可以参考其中一类的信息，或两类都考虑，亦即综合《消费者报告》杂志的调查分析和两名大学生有关他们偏爱的叙述。尽管先前只有不到15%的相类似的被试愿意接受两名学生之偏爱的影响，但是这一次，77%的被试都选择两类信息都要。很明显地，即使被试不认为他们应当受见证叙述的影响，但是他们最终还是难以抵挡见证信息的诱惑。并且，他们也的确受到了见证信息的影响。当被试选择同时看两类信息，而两类信息所推荐的避孕套又不同时，31%的被试选择了大学生见证叙述中所推荐的那种，而不是《消费者报

告》所推荐的。

另一个人们对鲜活轶闻做出特别反应的例证，是关于上世纪60年代中后期，媒体对越南战争的报道。随着战争没完没了的持续，士兵死亡名单仿佛永无尽头地增长，媒体不得不采取每周汇报上周战死的美军人数。周复一周，这个数字大致总在200～300人之间，渐渐公众似乎也习以为常了。然而，有一周，一个重要杂志用连续几页的篇幅，将上一周战死人员的照片全部刊登出来。这时公众非常具体地看到的是，大约250位活生生的人在过去一周内死去。结果，对这场战争所造成的损失之强烈抗议爆发了。250张照片所产生的影响是每周数字所远远不及的。但是作为一个社会的成员，我们应该克服这种不相信数字，而必须亲眼目睹才相信的习惯。绝大多数影响我们社会的复杂因素都是只有靠数字才能捕捉的。只有当公众学会重视以数字形式表达的抽象材料，如同重视形象表达的材料一样时，公众自己的观点才不会随屏幕上闪过的最新形象而左右摇摆。

不仅公众跳不出鲜活性这一因素的干扰，经验丰富的临床心理学家和医生在做出判断时，也总要尽量抵制单一个案的压倒性影响。作家法兰辛·拉索（Francine Russo）曾描述了弗吉尼亚大学肿瘤学家威利·安德森（Willie Anderson）所陷入的两难困境（Russo，1999）。安德森一直提倡用控制实验法，并且时时招募及储备一些病人来做有控制的临床测试。但是他知道在面对有些个别突出的病例时，情感因素会影响他的判断。这时，他不得不与自己的这种对单一个案的情感反应作斗争。尽管自己力求从科学角度出发，他还是承

认“当一个真实的人眼巴巴地望着你时，你就会被他们的希望，以及自己对他们希望的希望所围困，这的确是一个很困难的时刻”（p.36）。但安德森知道，有时对病人做的最好的事，就是不去看“望着你的那个真实的人”，而是根据最好的证据来做事。而所谓最好的证据是来自有控制的临床实验（将在第六章中表述），而不是对眼巴巴望着你的那个人所做出的情感反应。

评价证据时，鲜活性问题是我们要面对的一个难题。在到处都渗透着媒体形象的环境中，要整个社会不受形象的影响，而只依据可靠的证据来解决社会问题是越来越难了。作家贝瑞·格拉斯纳（Barry Glassner）曾讲述过一个大家都非常熟悉的例子（Glassner，1999）。1995年在“温佛瑞秀”的一次节目中，讨论是围绕着某一外科手术而展开的（为了不替鲜活效应助纣为虐，这里不点出这种外科手术是什么），这种手术曾经引起了一些争论，因为有人声称曾被它所伤害。但是来自梅奥临床医院，哈佛大学和密西根大学的研究证据都显示：这种手术程序总体上没有危险性——然而，正在这时，一名声称受到过这个手术伤害的妇女突然从观众席中跳出来，大叫道“我们就是证据。应该研究的是我们这些坐在这里的受害者”（Glassner，1999，p.164）。你认为数百万电视观众会对什么东西记忆犹新呢？——是梅奥临床医院的研究呢？还是大叫着说自己受过伤害的妇女呢？

总之，由于依赖见证证据而产生的问题经常存在。这种证据的鲜活性经常掩盖了更可靠的信息，从而混淆视听。但是心理学教师们都担心，仅仅指出依赖见证证据的逻辑错误还是不

够的，它不足以令学生们深刻理解这类数据所造成的陷阱。那么，我们还能做些什么呢？还有什么其他方式可以让公众了解这一概念呢？所幸倒真还有一种方法—— 一种略略不同于学术研究的方法，可以达到我们的目的。这一方法的重点就是用鲜活性来对抗鲜活性。让见证叙述搬起石头来砸自己的脚。让见证方法自身的荒谬来毁灭它自己。这一方法的施用者之一，就是我们独一无二、毋庸置疑的“神奇兰迪”。

神奇兰迪：以其人之道治其人

詹姆斯·兰迪（James Randi）是一名魔术师和一个什么都会的人，曾经获得过麦克阿瑟基金会所颁发的一项“天才”研究奖助金。多年来，他一直致力于教育公众一些进行批判思考所需的基本功夫。神奇兰迪（他的艺名）是用拆穿许多自称有通灵能力之人的骗术，来教育群众的。虽然他曾经揭发过无数假称自己有通灵能力的魔术师及大法师，不过最有名的还是在20世纪70年代他揭露通灵超级巨星尤瑞·格勒（Uri Geller）诈术的那一次。格勒以即席表演的方式大事鼓吹他自己的通灵力量，将全世界的媒体之焦点全部吸引过来了。几大洲的报纸、电视和主要新闻杂志上无不针对他的奇术争相报道。兰迪则察觉并披露出，格勒用以表演其通灵“绝技”的秘诀竟然是一些非常普通，以及一些简单的让人不好意思说出口的魔术小把戏，包括把钥匙或汤匙弯曲，启动钟表等等。对一个优秀的魔术师来说，这些是再平凡不过的手艺了。在格勒的骗术被拆穿并不停地被曝光之后，兰迪继续不断地用他那非凡的才智，揭露了像超感官感知、生物

节律、通灵外科手术、外星人、空中漂浮及其他各式各样的伪科学，从而维护了大众的知情权（Randi，1986，1987，1995；Sagan，1996）。

兰迪的许多较次要兴趣之一是向公众展示，任何荒谬的事情或无中生有的说法，都非常容易地可以收集到见证证据（Sagan，1996）。他的方法是让人们落入他们自己用见证所设下的陷阱中。兰迪充分利用了美妙的美国文化窗口——脱口秀，经常隐藏自己的真实姓名，假扮他人出现在这类节目中作嘉宾。在几年前纽约的一次节目上，他告诉观众，他那天早上驾车经过新泽西州时，看到一个橙色的V形物体从北面飞过头顶。几秒钟内，兰迪说:“电视台的电话接收机灯就像一个电子圣诞树一样的全亮了”。节目播出后，不断有目击者打电话进来证实他们也看到了这一奇异景象，并补充了许多兰迪在节目中“忽略”了的细节，其中包括：其实当时有不只一个“飞碟”经过，而是很多个。他们当然都没有想到，原来该“奇象”只是兰迪编出来的东西。这个小把戏一下子就说明了个人“发生了某一件事” 的经历叙述是完全靠不住的。

还有一次在加拿大温尼佩克，兰迪以“天文星相学家”的身份出现在一个电台节目上。一周前节目主持人曾邀请听众将他们的笔迹和出生日期提供给兰迪，好让他从中选出3位，并与他们保持联系，以便能够从他们的“生辰八字”中帮他们看看他们的性格会是怎么样的。他取得了巨大成功。在一个1～10分准确度量表上，这3位听众对他的预测的准确度评分是9、10和10。兰迪最终还是向听众们抖出了他的预测秘诀。原来，他只不过是一字不露地按照星相学家塞尼·奥摩

尔（Sydney Omarr）不久之前在一个电视节目中给3位观众所作的“预测”之原文，重读了一遍而已。

在另外的一次电台节目上，兰迪演示了另一种非常盛行的伪科学：生物节律（Hines，1998）。一位听众同意每天记日记，以供将来与一个为她准备的、为期2个月的生物节律图进行对照。2个月后，那位妇女打电话回来告诉听众：应该认真对待生物节律，因为她的图表准确地预测了她的实际行止，正确率达90%以上。兰迪不得不告诉她，他的秘书犯了一个愚蠢的错误，把另外一个人的图表给了她。之后，那位妇女同意看一看她自己真实的图表是怎么样的。于是一份图表很快又寄了给她，并要她再打电话来。几天后，那位妇女打电话回来了，说她松了一口气。原来，真正属于她自己的图表也相当准确。甚至，比原来那一张错的更准确！殊不知，在接下来的一次节目中，兰迪发现又搞错了，第二次送给那位妇女的竟然是兰迪秘书的图表，还不是她的！

兰迪的生物节律和天文星相学家诡计，实际上都是被称为巴楠姆效应（P. T. Barnum effect）的一类现象的例子（巴楠姆是著名的狂欢节目和马戏团的团主，他曾发明了这样一句话：“每一秒就有一个容易上当受骗的人出生”）。对这一效应，心理学家曾进行了深入的研究（如Dickson & Kelly，1985）。他们发现，绝大多数人认为泛化的人格特征描述都是正确的，并把这些特征视为是自己所独有的。这一巴楠姆效应令制造见证式证据变得非常容易，这当然正是说明了为什么这类证据是毫无价值的。心理学家的研究发现有一些词语及句子是绝大多数人都会认为适用于自己身上的，任何人都可以

把这些词语及句子贴在他们的"病人"或"顾客"身上，算是对他们个人"个别的"心理"分析"。而这些"病人"或"顾客"都会被这些个别化的人格分析之"准确性"所惊呆。他们完全不知道其实给予每个人的词语及句子都是一样的。当然，巴楠姆效应是在相信手相和占星术的人身上才会奏效（Kelly，1997、1998）。

兰迪在这些小诡计中所要达到的目的，是想给公众上一堂课，教育他们见证证据是没有价值的。他毫不例外地演示出，不管一个人说的是不是真的，如果他想制造出一些见证证据来证明自己说的是真的，那是太容易不过的事了（Randi，1983）。由于这一原因，通过见证来支持自己所提出来的特定理论或说法是毫无意义的。只有由在有控制的条件下所做的观察中获取的证据（在第六章里介绍）才足以检验这些理论及说法。

见证为伪科学打开方便之门

有时，有人会说，类似刚才所讨论的种种伪科学只不过是人们点缀他们生活的一种方式，大家笑一笑，并不伤大雅。老实说，谁会对这些事认真呢？不就是有几个人在异想天开，而另外几个人从中赚几个臭钱吗？

然而，如果我们对这一问题进行全面考察，就不难发现：伪科学盛行对社会的危害比人们想像的要大得多。第一，相信伪科学会导致个人悲剧。让我们来看看丽亚·萨利斯（Rhea Sullins）的个案。萨利斯是美国自然保健协会主席

的7岁女儿。这个协会主张人们可以用吃水果和蔬菜来取代用药和传统治疗。萨利斯生病后，“她父亲让她禁食18天，其间只给她喝水。接下来的17天里则只准她喝少量果汁”(Gilovich，1991，p.6)。可悲的是，萨利斯最终死于营养不良。更不幸的是，像萨利斯这样的例子比比皆是(见Kaminer，1999；Shermer，1997；Stanovich，2004；Sternberg，2002)。

在一个复杂、高科技的社会里，伪科学的影响可能由于某一些涉及到千千万万人利益的各种决策而得以迅速蔓延。也就是说，即使你不认同伪科学的理念，你也还是可能受到它的影响。例如，有些大银行和几家排名五百强的大企业，雇佣笔迹分析学家来帮助他们挑选员工(Sutherland，1992)，尽管有大量证据显示笔迹学对人才选拔毫无用处(Ben-Shakhar，Bar-Hillel，Blui，Ben-Abba & Flug，1989；Neter & Ben-Shakhar，1989)。从这个角度来看，伪科学的笔迹学指标令雇主忽视了其他更有效的选拔标准，结果导致经济上的无效率和对个别人士的不公平待遇。如果你因为你的笔迹中有一个很特别的小圆圈，而没有得到你想要的一份工作，你会怎么想？或者是因为一个通灵术专家看到你的“气场”出现一些乱流，而拒绝雇用你，你的感受又会是怎么样的？实际上，这确实发生在一些人身上。有些公司会付钱为申请工作的人进行“通灵分析”。例如，苏姗·金(Susan King)就是一个所谓有超感官感知、可以预见未来的人，公司付钱给她对拟聘用的人进行“预测”，来帮助决定要不要雇用他们。她说她不用亲自会见应聘者，就可以从照片和应聘者的

名字上看到她要的资料——虽然她的雇主有时会在初选名单出来后，把她请来参与最后面试，甚至面试后的鸡尾酒会，借以给她接近应聘者的机会（Kersham，1991）。在竞争激烈的经济时代，你想要他人用这种方式来决定你在哪里工作吗？

不幸地是，这样的例子并不少见（Shermer，1997；Stanovich，2004；Sternberg，2002）。当伪科学充斥社会时，它们会以多种方式影响我们——即使你不认同这种理念。例如，警察会雇请通灵人士帮助他们调查案件（Marshall，1980），即使研究证实这种做法并没有增加破案的几率（Hines，1988；Rowe，1993）。美国电视台（ABC）的节目制作人请好莱坞通灵人士帮他们决定节目的内容。试想，电视是美国社会最具影响力的电讯技术，竟然由通灵人士来决定节目内容（Auletta，1992，P.114）。最出乎我们意料之外的是，在里根总统主政时，白宫请了一个“星相学家”，来占卦决定“总统演讲、出席与各州州长见面、飞行旅程安排的时间，甚至还包括要讨论的议题”（Johnson，1991，p.454）。

像星相学这样的伪科学已经是一个很大的企业，涉及到报纸专栏、电台节目、图书出版、互联网、杂志文章以及其他传播工具。发行量最大的是占星术杂志，其销售量远远大于许多正当的科学杂志。美国众议院有关老龄化问题的事务委员会估计，全国每年人们被医学骗子骗去的钱超过几十亿元。总之，伪科学是一门相当赚钱的大生意，数以千计人的生计是靠公众对伪科学的盲目接受来维系的。

医学领域如何对待伪科学这一问题很值得我们心理学借

镜。因为在抨击伪科学，以及把正当的与不正当的医学实践划清界限方面，医学界的各类专业学会都比心理学界做得好。下面让我们看看由关节炎基金会出版，曾被美国众议院老龄化问题事务委员会所引述的一套识别不道德药品推销员的指南：

1. 他是以提供“特别的”或“秘密的”处方或仪器来“根治”关节炎。
2. 他刊登广告，在广告中用的手法是让取得满意疗效的“病人”讲述他们的“病例史”和见证经历。
3. 他保证（或暗示）可以很快或很容易就把病治愈。
4. 他可能会说知道关节炎的病因，进而说要清洗你体内的“毒素”，“增强”你的健康。他可能还会说，不需外科手术、X 线和医生的处方药。
5. 他可能会指责“正规医疗机构”过于谨慎，妨碍医学进步或是在迫害他……但他不会让他人用实验和实征的方法来检验他的方法（美国国会，1984，p.12）。

上面所列的这几条其实也可以用来作为识别带有欺骗性的心理治疗和理论之指南。当然，以上指南中的第 2 条涉及本章的主题。但是也要注意，第 1 条和第 5 条说明了在前几章所讨论过的内容之重要性：科学是公开的。除了用“见证叙述”作为“证据”之外，伪科学的从业人员经常试图逃避“公开地被证实”这一科学的标准，并指责他人是有意要压制他们所获取的“知识”。这样，他们就有借口带着他们的“研究成果”直接走进媒体，而不是把这些成果通过正规的科学出版程序来公诸于世。这种不正当的手段在心理学领域中

比在其他领域中用得更成功。因为与其他学科相比，媒体通常看不起正规的心理学研究机制。请大家注意这一点（第十二章将进一步讨论这个问题）。如果是物理学里的理论是未经证实的，媒体通常不会去报道，但是未经证实的心理学理论通常会被随随便便地报道出来，这是托伪科学经常在媒体被炒作之福，令新闻记者以为：心理学中“做什么都行，没有任何章法”。因此，心理学知识的消费者必须要警觉到：只要能吸引被试，电视和印刷媒体会大事报道心理学领域中任何稀奇古怪的说法，而不管这种说法与现有的证据是否相互矛盾。如此一来，受害者是一般老百姓。

各种神医神药的宣传使人对自己的病症产生一种过高的期望，这种期望在遭遇挫折时会给人们带来心理创伤。在我的资料中，有关这方面最卑鄙的一个实例，是一篇关于药店销售药品的文章，标题是“通灵人士显示盲人如何能用超感知觉来看东西”。人们经常因为迷恋上了伪科学，没能充分利用就围绕在他们身边的真正知识。通灵术的支持者等于是在鼓励人们在假药上乱花钱，并让患者忽视了真正有可能帮助他们的传统“非通灵”医学治疗（Angell & Kassirer，1998）。记者里察·布伦尼曼（Richard Brenneman）描述了通灵外科医生如何骗光了癌症患者的钱，并延误了本可通过合理医学手段进行的治疗（Brenneman，1990）。类似地，在一个有关医疗保健欺骗行为的市民大会上，密执安州反健康欺骗委员会的主席展示了一则虚假癌症治疗的宣传册子，上面宣称“本产品不能与其他癌症治疗方法同时使用”（“If It Sounds”，1990）。类似案例已造成的损害是无法估算的。

医生们越来越关注医学骗局在互联网上的蔓延，以及人类健康为此所付出的代价。马克斯 · 科普斯（Max Coppes）医生曾因气愤而致信给《新英格兰医学杂志》，警告大家医学骗子给人类所造成的真正伤害（Scott，1999）。他讲了一个9岁女孩的故事。癌症手术后，如果她接受化疗，将有50%的几率再活3年。然而她父母发现了一种未经验证有效的治疗法，这种方法用的是鲨鱼软骨组织来治病。他们采用了这种治疗法，结果那个小女孩只活了4个月就死了。

因此，任由伪科学蔓延的代价是巨大的。伪科学蔓延的主要原因就在于，人们混淆了哪一类证据可信及哪一类不可以用来增加某个说法的可信度。由于见证叙述能很方便地为几乎所有理论及说法提供支持，并通过这些支持造成巨大的影响，见证型证据就为伪科学的发展和获得信任敞开了一扇大门。对心理学知识的消费者来说，除了小心提防外，没有其他更好的招术可施。在接下来几章里，将讨论要证实一个理论或说法，到底需要哪一类型的证据。

小　结

个案研究和见证叙述在心理学（及其他社会科学）的研究初期是有用的。因为那时我们去发现可以进行研究的现象和变量是至关重要的。虽然个案研究在早期的、理论形成前的科学研究阶段是有用的，但是在研究的后期，亦即对提出的理论进行特定检验时，它就变得毫无用处了。在心理学中，个案研究和见证叙述对理论检验没有价值的一个主要原因是，

有安慰剂效应在作祟。安慰剂效应是指人们有报告任何治疗对他们都有帮助的倾向，不管这种治疗是不是真的有疗效。安慰剂效应的存在使我们不可以采用疗效见证来作为支持一个心理（或医学）治疗的有效性。原因是安慰剂效应保证：不管什么治疗方法，都有可能获得有关它的疗效的见证资料。

尽管见证证据对理论检验毫无用处，心理学研究还是发现：这类证据由于鲜活性效应而被人们格外看重。鲜活性效应是指：人们过分看重更鲜活和更容易从记忆中提取出来的证据。对大多数人来说，见证证据就是一种极具鲜活性的信息。这带来的后果是：人们在验证某一心理学的理论说法时，往往过分依赖这类证据。实际上，见证叙述和个案研究型证据根本不能用来支持任何理论或说法。

（肖崇好　译）

第五章　相关和因果

——烤面包机与避孕

几年前，台湾曾做过一个大规模的调查，研究哪些因素与人们是否使用任何避孕方法有关。由一些社会学家和医学工作者组成的庞大研究队伍，广泛收集了诸多行为和环境变量的数据。研究者想知道什么变量能最有效地预测人们是否使用任何方法进行避孕。收集数据后，他们发现人们是否使用任何避孕方法与一个变量的相关最高，那个变量就是“家中拥有家用电器（如烤面包机、风扇等）的数量”（Li，1975）。

这个研究结果恐怕不会让你想要用在各高中免费发放烤面包机来解决青少年的怀孕问题吧！但是，你为什么又不会这样想呢？既然小家电和是否使用任何避孕方法之间似乎存在着密切的关系，并且这个变量又是所有当时测量的众多变量中最具预测性的一个，你应该没有理由不这样想。我希望你对这个问题的回答能是：在这个研究结果中，是这两个变量关系的“性质”，而不是其“强度”与研究者想理解的问题相关。试想一下，如果我们举办一个“在高中免费发放烤面包机”的活动，人们将会怎么想呢？大家就会以为烤面包机导致了人们去使用任何避孕方法！这不是很荒谬吗？至少在上面所举的这个“显而易见”的例子中，我们很容易便能看出这两个变量之间只可能存在“关联”关系，但决不可能存在“因果”关系。

在这个例子中，我们不难猜到，使用任何避孕方法和拥有小家电数量之间的联系，可能是因为这两个变量同时与一个第三变量都有关联所产生的结果。例如“教育水平”可能就是一个中介变量，我们知道一个人的教育水平与其是否使用任何避孕方法，以及与其“社会经济地位”都有关联，而社会经济地位高的家庭又倾向于拥有比较多的家用电器。当然，这两者之间可能还有其他更多的变量在起作用，我们在此仅想说明，无论烤面包机的数量与使用任何避孕方法有多大的相关，它们之间也不会意味着有因果关系。

这个例子应该很容易会让我们理解本章所要谈的基本原则：有相关并不表示一定有因果关系。本章中将要讨论的是令我们无法做出因果推论的两大问题：第三变量问题与方向性问题。在讨论完这两个问题之后，还将讨论选择偏见如何会经常引起第三变量这一问题。

“相关”作为研究证据的局限性有时是很难察觉的。而且我们很容易就会将相关看成是因果关系。常见的情况有：当我们以为他们一定存在因果关系时，或当我们心存偏见时，或者当我们一心只想要寻找证据来支持自己的理论时等等。卡尔·皮尔逊（Karl Pearson）的个案提供了一个很清楚、但颇具讽刺意味的例子。

第三变量问题：古德伯格和糙皮病

20世纪初，数以千计的美国南部人遭受了一种皮肤病（糙皮病，*pellagra*）的侵袭，每年约有100 000人致命。其

症状有：头昏眼花、昏睡无力、伤口不愈、呕吐、严重腹泻。这种病被认为是由某种“来历不明”的微生物感染所致（Chase，1977，p.205），当时有研究资料显示，这一病症与卫生条件有联系。这一结果，令全国糙皮病研究学会的医生都认为是很有道理的，因为在南卡罗来纳州的斯巴达堡市内没有受糙皮病干扰的人住的房子似乎都有屋内自来水及污水管道系统。也就是说，卫生条件较好。这种相关恰好验证了当时对传染病病理的一个理论，该理论认为由于卫生条件较差，污水处理不当，糙皮病是通过感染者的粪便不断传播开来的。

一位叫约瑟夫·古德伯格（Joseph Goldberger）的医生对这种解释不以为然，他在美国公共卫生部部长的指示下对糙皮病进行了好几项调查研究。古德伯格认为糙皮病是由于饮食营养不良所引起的，因为美国南方地区普遍比较贫穷，人们主要依靠谷物、燕麦和玉米粥这样的高碳水化合物的食物为生，而少吃肉、蛋、奶等蛋白质含量较高的食物。古德伯格认为污浊的环境和糙皮病之间的相关关系并非因果关系（就像烤面包机与避孕行为的例子），之所以会出现相关是因为备有下水设施的家庭通常经济条件也比较好，而这种经济差别也反映在他们日常的饮食中，例如是不是食用较多的动物蛋白。

但是，好像有些不对劲！为什么我们要相信古德伯格的“营养缺乏”理论而不相信前面的“粪便传播”理论呢？毕竟，两派人马都是坐在那里根据相关数据在推论什么才是造成糙皮病的原因。为什么医学会的医生不能说古德伯格的相关同样是在误导大家呢？古德伯格又有什么理由排除“粪便处理不当”这一理论假设呢？古德伯格之所以理直气壮是因为他做了

一件我刚才忘了告诉你们的事：他曾亲自吃下糙皮病患者的粪便！

为什么古德伯格的证据比较好

支持古德伯格理论的证据不仅只包括靠观察变量间相关系数所做的推断，还包括了另一类的研究证据。那就是对关键变量进行了操控之后所收集的数据（有关操纵控制，将在下一章再进一步讨论）。这一方法通常涉及布置下一个在自然情况下极少出现的特殊场景。为了要制造出这样一个特殊情景，我们说古德伯格是想出了一个相当“不自然的”方法，是绝对言不为过的。

为了证实糙皮病不会传染，不是通过患者的体液传播出去，古德伯格在自己的身上注射了患者的血液，还将患者喉中和鼻中的分泌物塞进自己的口中。除此之外：

> 最后，他选出了两位糙皮症病人，一个有皮癣症状，一个有腹泻症状。他从病人身上拔下鱼鳞状角质皮混入4毫升该病人的尿液，并加入4毫升他们的粪便，然后与4小撮面粉揉在一起做成小药丸。他自己及几个助手，还有他的妻子，自愿吞服了下去。（Bronfenbrenner & Mahoney，1975，p.11）

古德伯格和其他志愿吞食的人都没有患糙皮病。这就说明，他已经布置下了这种病如果是通过感染传播的所有可能的途径，结果什么事也没发生。

古德伯格设想并布置了几套操作程序，来控制别人认为是糙皮病“因果机制”的变量，从而证明了这一机制对该疾病

的无效性。接下来，他的任务是证明自己的因果机制是有效的。他从密西西比州监狱农场找到了两批无糙皮病并自愿参加实验的犯人，其中一批配给高碳水化合物/低蛋白的食物，这是他预期会导致糙皮病的原因；而另一批人则配给相对营养均衡的食物。5个月内，食用含高碳水化合物/低蛋白食物的犯人们果然为糙皮病所肆虐，而另一批人则安然无恙。尽管后来古德伯格的观点被一些因政治目的而企图否定美国存在“贫穷人口”（到吃不上蛋白质食品的地步）现象的人强烈反对，但经过一段长时间的被打压，最终由于他的理论确实比其他人的得到更多实征数据的支持，还是得到了认可。

糙皮病的这段历史告诉我们，如果是依据相关研究的基础来制定社会和经济政策，必将使人类付出惨痛的代价。但这并不是说我们最好永远都不要使用相关数据为研究证据。恰恰相反，在许多情况下（例如：当我们只想作预测，不探究其原因时），它正是我们所需要依靠的证据（见第八章），科学家经常必须在没有充分的信息及知识的情况下，去解决难题。重要的是我们应该在运用相关数据时要小心谨慎。像“糙皮病—污水”一案例在心理学领域中也经常发生，这也就是第三变量的问题。“糙皮病—污水”案例中的两个变量的相关并不能说明它们之间有某种直接的因果关系，这一事实说明的很可能是与这两变量都相关的第三个变量造成了两者的相关，在该例中这第三个变量是“饮食营养”。糙皮病和污水的这种相关通常被称为是虚假相关，因为两者所出现的联系并不是因为所关注的这两个变量之间存在因果联系，而是由于这两个变量均与第三变量相联系所致。

再来看一个发生在现实生活的例子。公立学校与私立学校教学质量的比较一直是人们争论的话题。在这个争论中所下的一些结论可以说是相当鲜活地展示了从相关数据推断为因果关系的弊端。私立学校比公立学校好还是坏，应该是一个可以用实证研究可以解决的问题，可以通过社会科学中的调查法来辨真伪。但是，这并不表示这是一个很简单的问题，只是说明它是一个科学问题，有被解决的可能性。所有倡导私立学校比较好的人都对这一点心知肚明，因为在他们维护自己的观点时，常常用这样的一个实征研究的结果：私立学校学生的成绩比公立学校的要高。这一事实并无可争辩，各类研究和教育统计数据在这方面都得到很明显、也很稳定的结论。但是问题在于，仅用这些学生的成绩是否真的就可以下结论说：是私立学校的教育本身导致他们学生成绩比较好呢？

教育效果测验的结果是与许多因素相关联的，而这些因素彼此之间又都相互有关联。要评估公立与私立学校的好坏，我们不能仅用学校类型与学生成绩之相关值这一个指标。例如学生成绩与家庭环境因素，如父母的教育程度、父母所从事的职业、该家庭的社会经济地位、家中藏书数量多少等都有关系。这些因素与孩子能否进入私立学校的可能性又存在相关。因此，家庭环境作为一个潜在的第三变量可能是导致学生成绩与学校类型相关的原因。总之，学生成绩可能与学校质量无关，家庭经济条件好、有良好教养的孩子，本来就更有可能被送到私立学校去读书。这一类学生即使在公立学校上学，其学业成绩也会很好。

很幸运的是，现在我们有了许多像多元回归、路径分析

这类复杂的相关统计方法（顺便提一下，统计学的发展许多可是心理学家的功劳），可以用来帮忙我们理清这些问题。这些统计方法可以计算出两个变量在排除了其他变量干扰之后，比较“纯净”的相关值。杜克大学的佩奇和基思（Page & Keith，1981）曾用这种复杂的相关技术分析了在美国全国教育统计中心（National Center for Educational statistics）的帮助下所收集的有关学生情况的统计资料。他们发现，排除一些影响学生家庭环境和一般性智力因素之后，学生学业成绩与学校类型之间不存在任何相关。他们的这一结果也得到其他研究者的确认（Berliner & Biddle，1995；Wolfle，1987）。

因此，提倡开办私立学校以加速教育发展，和想用烤面包机来控制生育，看来是没什么两样。学生成绩与就读学校相关不是因为它们有直接的因果关系，而是因为家庭环境以及智力水平的差别罢了。

复杂相关统计方法可以帮助我们排除第三变量的影响，但是这并不表示这个排除过程一定削弱原有相关的强度。有时候，当我们排除了第三变量之后，两变量之间相关的强度仍保持不变。这时，这一相关就告诉我们许多东西了。很可能，两者的相关并非由第三因素所造成，而是的确存在。当然，也有可能是还有一些没有想到的，或无法排除的其他变量造成了虚假相关。

下面所述托马斯（Thomas）、亚历山大（Alexander）和埃克兰德（Eckland）的研究（1979），也提供了一个很好的例子。他们通过调查发现：高中生能否上大学与他们家庭的社会经济地位相关。这个重要发现，无疑击中了美国这

样一个以“能力”为价值核心的社会的要害，因而引起很大的反响。因为这一研究结果不正是在暗示，一个人一生会不会成功不是取决于能力和努力，而是取决于他的社会经济地位吗？然而，在我们下这个结论之前，也许应该先考虑一些可能用以解释这一数据的其他假设。也就是说，上大学与社会经济地位之间的相关很可能只是一种假象。很容易想到的一个可能会对两者都有影响的因素是学生的学业能力，它与读大学以及社会经济地位两者都有关系。如果这个因素被考虑进去，那么原来那两个变量之间的相关或许就不存在了。但当调查人员在处理了这一第三因素后，再重新对两者进行统计分析时，却发现即便排除了这一第三因素，上大学与社会经济地位之间仍然存有显著相关。因此，看来高经济收入家庭的孩子更容易考上大学并不完全是由于他们的学业能力比较强所决定的。正如前面所说，这个结果当然并不能排除其他第三变量造成两者相关的可能性，但是能够用这样一个重新分析来排除学业能力对两者相关的影响，本身无疑也在理论及实践方面具有很显然的重大意义。

两位安德森教授就曾经通过检验其他假设能否解释同一组研究数据的方法，验证了一个关于暴力具有地区差异的理论（Anderson & Anderson，1996）。他们所使用的方法是偏相关。曾经有研究数据显示美国南部的暴力犯罪率要高于北部各州。两位安德森对这个高温假设（heat hypothesis）进行了验证，这一假设认为南部地区所以有较高的暴力事件是因为“令人不舒适的高温导致了人们的攻击性动机和（有时候）攻击行为的增多”（p.740）。找到一个城市平均温度与其犯

罪率之间有相关，并不稀奇，这很容易做到，许多人都报告了这一结果。然而，当他们发现即使将像失业率、人均收入、贫困率、教育状况、人口数量、人口平均年龄及其他几个可以是第三变量的因素都加以控制之后，温度与暴力犯罪之间的相关仍然显著。这就使得高温假设理论的可信度大大提高了。

方向性问题

如果我们可以通过操纵变量来顺理成章地推断因果关系，就没有理由要像上面说的那样转弯抹角地靠分析相关系数来推论。但是很让人苦恼的是，一旦涉及到心理学的问题，我们经常都必须靠相关系数来做进一步的分析。并且在心理学知识对解决社会问题越发重要的今天，这种依赖相关分析的趋势越加严重。教育心理学里一个广为人知的例子可以很好地说明这一点。

人类对阅读进行科学研究大概始于100年前。从那时起，研究者就发现眼动型式（eye movement patterns）与阅读能力之间存在相关，阅读能力差的人在阅读文字时眼动没有规律，并经常回扫（从右到左）和凝视（停下不动）。基于这一相关，一些教育学家认为缺乏眼动技巧是导致阅读困难的原因。于是，许多眼动训练项目开始开展，并且很快地应用于小学的教学中。这些教学项目都是在还没有弄清楚这两个有相关的变量到底哪一个是因，哪一个是果的情况下，就早已被激活施用了。

而今天我们已经很清楚，人们从这一个相关所推论出的，

由眼动到阅读能力的因果关系，恰恰与实际情况呈相反的方向。无规律的眼动不会导致阅读困难（Olson & Forsberg, 1993；Rayner, 1998）；而是阅读速度慢、理解困难导致了无规律的眼动！小孩子在识字水平和理解力提高的过程中，眼动会逐步趋于有规律。相反地，训练孩子的眼动不会提高他们的阅读水平。

最近十几年，研究者已经清楚地发现，词语解码及语音加工时所出现的语言问题才是阅读困难的根源所在（Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001, 2002; Shaywitz, 1996; Stanovich, 2000），而很少发现是由于不良的眼动型式导致阅读能力丧失的例子（Olson & Forsberg, 1993, Share & Stanovich, 1995）。但是如果你现在到那些中等或以上规模的学校地下储藏室里去仔细找一找的话，你一定会找到一个盖满灰尘的眼动训练仪，到如今，这些仪器已成为浪费公物的样板。这全是由于人们习惯把一个相关当作是因果关系所造成的恶果。

另一个有些相似的例子曾在第一章中讨论过。在教育社会工作领域里，曾一度流行认为学业成绩低、滥用毒品、少女怀孕及其他许多问题行为是由低自尊造成的（Dawes, 1994; Kahne, 1996）。这个因果关系的方向似乎是理所当然的：低自尊导致问题行为；高自尊导致优秀的学业成绩及其他成就。因此在这一因果推论的激发下，各类提升自尊的训练项目应运而生。但是这里存在的问题与“眼动”的例子一样：这个因果关系的结论仅仅是从两个变量的相关关系中推衍出来的（并且，在这里相关值非常小，请见 Kahne, 1996）。如今

已经得到证实，即便是自尊与学业成绩存在因果关系，其方向却极可能与前面的推论正相反：在学校（以及生活中其他方面）的优秀表现导致了高自尊（Baumeister，1999；Ruggiero，2000；Stout，2000）。

我们前面的讨论主要是围绕着两变量的相关常常涉及的两类问题：第一类叫做方向性问题，“眼动”和“自尊”的例子可以说明，当变量 A 与 B 有相关时，在下结论这一相关是由 A 导致 B 之前，必须想到这种因果关系的方向很可能是相反的。也就是说，很可能是由于 B 的变化而引起了 A 的变化。第二类问题是第三变量问题，糙皮病（和烤面包机与生育控制以及私立学校与学业成绩）的例子可以说明，两变量之间的相关或许不能告诉我们任何一方是另一方的因或果，很可能只是由于两个变量均与第三变量有相关而导致的。

选 择 偏 差

在有一些情况，虚假相关存在的可能性是很高的。那就是当选择偏差出现的可能性很高的时候。所谓选择偏差就是指：某些人与某些环境存在着一种关系，是由于具有某些特定生理、行为、心理特征的人会选择生活在具有某一特定的环境居住。因此，选择偏差造成环境特征与行为及生理特征之间的虚假相关。

下面通过一个例子来看看选择偏差是如何产生虚假相关的。请你很快地告诉我们美国哪一个州，由呼吸道疾病导致死亡的人数最多。这个问题的答案当然是亚利桑那州啰！为什

么呢？亚利桑那州的空气不是已经净化得很好吗？难道说洛杉矶的废气浓烟会扩散到那么远的亚利桑那州吗？凤凰城的郊区环境已经变得那么差了吗？不可能，绝对不可能！别急，让我们好好想想：或许亚利桑那州的空气也很清新，或许只是有呼吸道疾病的人喜欢移居到那里，然后就在那里去世。呀！这样想就对了。这让我们注意到，如果一不留意，就很可能被呼吸道疾病与死亡人数的相关所误导，以为是亚利桑那州的空气是致命的。

然而，选择偏差有时很难被发现，它们经常会被人们所忽略，特别是像自尊的例子那样，当我们心中早已存在一个特定方向的因果联系的想法时。这时，即使是再聪明的头脑，诱人的相关系数和心中的预设偏见也会联合起来把我们蒙骗。下面是一些具体的事例。

有关美国教育质量这一问题在过去20年里人们一直争论不休。在这一问题的辩论中，我们可以看到考虑选择因素的重要性。当时，大量的统计数据曾一股脑地涌向大众，但是却没有给他们提个醒儿，警告他们要避免由相关数据去推论因果关系，因为相关数据内含有大量的具有误导性的选择偏差。

在整个争论过程中，许多带有政治目的的教育评论家一再试图提供有关教育质量与教师工资水平及班级大小规模没有关系的证据，尽管已有许多研究表明后两个变量对前者都很重要（Ehrenberg，Brewer，Gamoran & Williams，2001；Finn & Achilles，1999）。他们提到的证据当中，有一个是对50个州的学生之学业能力测验（scholastic aptitude test，SAT）结果所做的分析。数据显示测验的平均分数的确与教师工资和教育

支出的联系很小。即使有联系也与人们所期望的方向正相反。一些教师平均工资高的州，学生测验成绩的平均分很低；而有些教师平均工资低的州，学生的测验得分却很高。但是，在这个研究中，只有那些想考大学的高中生才参与了这次测验，这就使我们看到选择偏差导致虚假相关是有多容易。

从测验中，我们看到密西西比州的学生在SAT（学业能力测验）上的得分要高于加利福尼亚州的学生（Powell & Steelman，1996；Taube & Linden，1989），并且差异还相当大，密州比加州的平均分要高出100分之多。由于密西西比州的教师工资是全国最低的。所以这就成了那些极力要求削减教师工资的保守派人士提供了一个庆祝的理由。但是，密西西比州的学校是否真的比加利福尼亚州的学校好呢？再者，是不是前者的整体教育状况都比后者好呢？当然不是。几乎所有其他的客观指标都显示加利福尼亚州的学校要更胜一筹（Powell & Steelman，1996）。如果真是如此，又如何解释SAT的结果呢？

答案就在于选择因素。不同于许多要求全州教育系统之所有学生都要参加的标准化测验，SAT由于不是所有高中生都一定要参加的测验，因而存在选择偏差（Powell & Steelman，1996；Taube & Linden，1989；Wainer，1989）。只有那些认为自己有希望进入大学的学生才参加这项测验。这个因素是导致了各州测验的平均分数存在差异的部分原因，同样也解释了为什么一些教育实力最强的州在学业能力测验中平均分偏低的现象。

选择因素以两种途径作用于各州的SAT分数。首先，一些州立大学入学的要求是看美国大学入学考试（American col-

lege testing，ACT）的分数，而不是SAT分数。因此，在这些州参加SAT考试的学生都是想到其他州读书的学生，而能够去外地读书的学生往往家庭条件比较好、学业能力也高于该州平均水平。这就是为什么会出现密西西比州与加利福尼亚州的差异情况：在密西西比州只有4%的高中生参加SAT，而加利弗尼亚州的数字则是47%（Powell & Steelman，1996）。

第二个选择因素则微妙得多。在一些教育系统发达的州，许多学生喜欢读完高中以后继续读大学深造，所以在这些州中，很高比例的学生包括相当一部分能力稍差的学生也都参加了SAT。而在一些辍学率高、教育质量较差的州，只有较少的学生想读大学，因此，这些州最终参加SAT的学生仅仅是最有能力上大学的那群人。这样一来，当然是这些州SAT的平均分要高于那些报考比例大的州。

错用SAT分数也提供了一个机会让我们看到，当公众缺乏本书所传授的基本方法和统计思想训练时，要想纠正误用统计是何等的困难。在1983年这本书的初版中，我就以SAT分数为例来解释选择偏差。10多年后，在1994年的第四版中我又讨论了一位印度教授鲍威尔（Powell）的一篇关于分析保守派政治专栏作家威尔（George Will）的文章。如你所料，Will在他的文章中声称由于高SAT分数的州教育投资并不高，所以他反对在教育中拨款太多（Will，1993）。然而，鲍威尔指出，那几个被单独列出的所谓分数特别高的州：爱阿华州、北达科他州、南达科他州、犹他州和明尼苏达州，其SAT的参加率分别仅为：5%、6%、7%、4%和10%；相反地，全美国有超过40%的高中毕业生参加了SAT测试。因为在这

些州中，公立学校所需的入学测试为ACT测试，只有那些计划去其他州学习，而且“通常是到声誉很好的私立学校”去读书的学生才参加SAT测试（Powell，1993，p. 352）。相反地，在被威尔点名为SAT得分低而教育消费高的新泽西州中，却有76%的高中毕业生参加SAT测试。显然，北达科他和南达科他州参加SAT测试的学生，与有3/4的学生参加了这项测试的新泽西州的学生相比是更是“拔尖”的一群。

在《教育研究者（*Educational Researcher*）》杂志上，心理测量学家霍华德·韦纳（Howard Wainer）曾分析了1993年6月22日发表在华尔街日报上的一篇文章。这篇文章介绍一个由美国传统基金会所做的研究。这个基金会是一个带有较强烈社会政治理念色彩的智囊团，反对在教育上投入太多的经费，他们的证据是，（对了，想必你已经猜到了）SAT分数较低的州在教育中的投资反而较多。好在，韦纳的文章不仅解释了是选择偏差导致了这一结果，它还让我们看到，当把一个测验（全国教育进展评估，或称NAEP）施于具有代表性的样本，而不是顺着自主选择所得到的样本时，这种关系的因果方向会反过来：在教育中投资越多的州，其平均得分越高。

运用前面提到的偏相关分析法，鲍威尔和斯蒂尔曼也证实了这种联系（Powell & Steelman，1996）。他们发现，当用统计学的方法将各州参加测试的学生比例纠正到相同时，那么每个州在每个学生身上增加1 000美元的教育投资，与其SAT平均分数增加15分，有着关联性。然而，尽管有大量证据表明如果不作统计纠正，选择偏差将使得各州之间的SAT

分数不具有可比性，但媒体和政客们为了他们的政治目的仍在继续使用这些未经修正的数字来误导大众。

在临床心理学上的一个事例可以表明选择偏差问题可以是多微妙和“邪恶”。研究数据经常表明，对于各类成瘾问题，如：肥胖、吸烟、吸毒等，做过心理治疗的人的治愈率要低于那些没做过心理治疗的人（Rzewnicki & Forgays，1987；Schachter，1982）。其原因，你会很高兴知道这一点，并不是心理治疗使成瘾行为更加难以改变，而是在于那些寻求心理治疗的人，其成瘾问题比较不容易被察觉，而且无法自愈的情况也较少。

韦纳（Wainer，1999）讲述了二次世界大战中的一个故事，也提醒我们要小心选择偏差。一个飞机分析师想通过分析返航后的飞机被子弹击中的洞孔分布，来确定在飞机上安放加固防弹层的位置。通过分析他认为应该将防弹层安置在返航机上没有被子弹袭击的位置。他的理由是，子弹应该是均匀地袭击了飞机的各个部位。所以，如果一架飞机能返回，就表示这架飞机被子弹击中的地方必定是不会对该飞机造成致命损伤的地方。那些没有子弹洞孔的地方，看来都是要害，因为被击中该处的飞机都没能返航。因此，防弹装置应该安装在返回飞机中没有被子弹击中的部位！

简而言之，本章推荐给“消费者”的原则很简单：注意选择偏差的可能性，并应避免仅仅从相关数据中去推论因果关系。当然，我们的确可以通过某些复杂的相关设计做出有限的因果推论；也可以对某些研究问题由于其结构的特殊性，而令相关数据用来作为检验因果假设的基础。同样地，相关

证据也有助于表明众多研究证据已经集中支持一个假设的现象（见第八章）。尽管如此，对于心理学知识的应用者来说，宁可犯太过挑剔的错误，也不要被相关数据所欺骗，误认为两个变量有因果关系。

小　结

本章的中心议题是想告诉大家，两个变量之间仅仅存在相关，并不能保证一个变量的改变会导致另一个变量的改变——相关并不意味着因果关系。在解释相关联系时，我们讨论了两个问题。首先，在第三变量问题上，两个变量之相关并不意味着它们之间有直接的因果关系，或许是由于一个尚未被测量的、而又与它们俩都有联系的第三变量所导致。实际上，如果潜在的第三变量被测量了，就能够用相关统计法，如偏相关（在第八章中会再讨论），来评估第三变量是否决定了二者之间的联系。第二个解释相关的困难，在于存在方向性问题：即使两变量间确实存在着直接因果关系，到底哪个是因，哪个是果，也无法仅从相关中反映出来。

选择偏差是行为科学中许多虚假相关产生的原因。所谓选择偏差是指：人们总是或多或少人为地选择自己的居住环境，因此造成某些行为特征与某些环境变量之间有相关。如本章古德伯格的例子所展示的，以及在下两章中会更加详细论述的：排除选择偏差的惟一方法就是在严格操控各变量的情况下来进行真正的实验研究。

（徐四华　译）

第六章　一切尽在掌控之中

——聪明汉斯的故事

本章的开始，让我们先来做一个测验。噢，别担心，不是测验你前几章读得好不好。其实这个小测验非常简单，它们与我们平时都见过的一些物体移动现象有关。请准备好一张纸，现在开始：第一题，设想一个人拿一根绳子在他或她的头顶上绕圈，绳子的远端系着一个球。从人的头顶的正上方，我们可以看到这个球运行的轨迹，是一个圆圈。请将这个圈画在纸上，并在这个圈的一处画一个点。然后，在这个点和圆心之间画一条直线，直线代表那根绳子，点代表在特定时刻的那个球。假设在某一旋转瞬间，绳子被剪断了，你的第一个测验任务就是用笔画出这个球飞出后的运行轨迹。

你的下一个测验任务是，设想你是一个轰炸机的飞行员，现在正以每小时500米的速度在6.096千米的高空冲向目标，为简单起见，假设没有空气阻力。那么你要在哪个点上投放炸弹，才能击中你的目标？在到达目标前？在目标的垂直上空？还是在飞越了目标之后？请说出你选择的投放点与目标间的具体距离。

最后，设想你正在进行枪与肩同高的步枪射击运动，假设没有空气阻力，并且枪与地面平行。如果一颗子弹从与枪相同的高度坠落的话，到地面需1.5秒。那么假设你现在由枪

管中射出一发子弹，其起始速度为每秒609.6米，它需要多长时间才能落到地面呢？

好了，测验完毕！答案？噢！对了，答案我们会在后面讲到。但是首先，为了了解我们对运行物体知觉的准确性，需要更全面地了解科学家们所用的实验方法背后的思考逻辑。本章将探讨实验控制和操纵（manipulation）的一些原理。

斯诺和霍乱

在前一章中我们讲到，约瑟夫·古德伯格对糙皮病的研究，一部分依靠的是来自他认为这种糙皮病是不会传染的灵感。然而，在20世纪70年代早期，约翰·斯诺（John Snow）在寻找霍乱病病因的研究时，则将病因放在相反的猜想上，最后也取得了成功（Goldstein & Goldstein，1978）。早在18世纪50年代，在伦敦，人们对不断暴发的霍乱疫症，提出许多理论，而且彼此争论不休。很多医生认为霍乱病人呼出的气体会导致感染，这就是所谓的“秽气理论”。但是斯诺的想法却是：疾病是通过被病人排泄物污染的供水系统传播出去的。

斯诺决定检验自己的想法。幸运的是，在伦敦有很多供水源，不同的供水源供给不同的地区。因此，如果他的构想是对的，霍乱的发生率应该因供水源受污染程度的不同而有差别。但是斯诺也想到了，这种比较可能会出现严重的选择偏差（请回想在第五章的讨论），由于伦敦的各市区贫富差别很大，因此，供水源与各地区的患病率之间的任何相关都会由于其他影响健康的，与该地区的经济发展水平相关的变量，

例如，饮食、压力、职业安全及衣食住行的质量等。总之，斯诺所得到的相关可以是一个虚假相关的可能性，和第五章中所讲述的糙皮病与污水系统的例子一样高。然而，斯诺非常机灵地注意到这个问题，而且利用了当时的一个特殊情境，解决了这一问题。

在伦敦的一个市区，恰巧有两个自来水公司，给当地居民提供饮水，而且是随机的，没有什么系统规则地供水。也就是，即便是紧邻的两个家庭可能是由不同的自来水公司供水。并且斯诺发现，这个地区内，人们的社会经济状况基本上非常接近。然而，即使在这么得天独厚的条件下，如果两家自来水公司的水都同时受到污染，斯诺还是无法验证其假设。幸好，斯诺是一福星，这两家公司的水并未同时受到污染。

在一次霍乱流行过后，兰姆博斯（Lambeth）自来水公司为了免受伦敦污水系统的污染，把公司迁到了泰晤士河的上游。而南沃克—沃克斯霍尔（Southwark and Vauxhall）公司却仍驻扎在该河的下游，这就使兰姆博斯公司的水受污染的可能性比南沃克—沃克斯霍尔公司的小很多。斯诺运用化学检验证实了这一点。下面是饮用不同供水公司所提供的水之家庭的霍乱死亡率：兰姆博斯公司为每 10 000 家庭中有 37 人死亡，而南沃克—沃克斯霍尔公司为每 10 000 家庭中有 315 人死亡。

在这一章中，我们将讨论在斯诺和古德伯格的故事中，到底隐含着什么科学思考的逻辑。这对我们理解科学家到底在干什么非常重要。如果不能理解这种逻辑，他们所做的事就很可能被看成是神秘的、古怪的、甚至是荒谬可笑的。

比较、控制和操纵

尽管市面上已经有很多厚重的书，讲授科学研究方法，但是对于从未做过实验的门外汉来说，这些书大概都是不必要的，因为他们只想知道一个大概，并不想被实验设计的所有复杂细节所羁绊。科学思考最重要的几个特点其实非常容易掌握，那就是比较、控制和操纵。为了要对一个现象有更基本的了解，一个科学家会对各种情况进行比较。如果没有比较，看到的都是一些孤立的观察或事件。就像第四章所叙述的那样，对这类观察的解释往往会是非常模糊的。

通过在不同条件下比较数据的异同，科学家可以推翻对数据的某些解释从而确定另外的一些。他们设计“比较”，来除一些同样可能进一步成为解释其研究结果的其他说法或理论。也就是说，我们要在一个实验设计中尽可能多地排除不正确的可能解释。要做到这一点，他们或者通过直接控制实验条件，或是通过观察允许进行比较各种可能解释的自然情境。

这后一种途径在霍乱这个例子中得到了很好的体现。那两家自来水公司并不是斯诺随便找来的。他意识到，虽然有许多地区都是由不同的自来水公司来供水，但是这些地区在与健康相关的许多社会经济指标上，也会存在很大的差异。因此，即使观察到霍乱在不同地区的不同发病率，也难逃要面临“许多不同解释可以同时存在”的难题。斯诺非常清楚，只有排除无关的解释才可以推动科学发展（请回想一下第二章所讨论的可证伪性），所以他努力去寻找，并且找到了一个

"比较"情境，在这个比较情境中，许多认为霍乱疫情是与社会经济水平相关的大量其他的可能解释，被排除出去了。

斯诺能找到这样一个可以让他消除其他解释的自然情境，是非常幸运的。这种在自然情况下产生的"比较"条件是不多见的。但是，如果我们因此以为科学家们都是守株待兔地在那里等待这种情境的出现，那就是非常荒谬的。事实上，大部分的科学家都是在尝试着去重构我们的世界，从而可以用之来分辨各种可能的假设，看看哪些假设是对的，哪些是错的。要做到这一点，就必须操纵那个认为有可能是原因的变量（如斯诺例子中的"供水污染"），并观察，当保持其他变量恒定时，是否会产生不同的结果（"霍乱发病率"）。

因此，一个好的实验设计应该是，由科学家去操纵他感兴趣的变量，并控制其他可能造成影响的无关变量。需要注意的是，斯诺并不是这样做的，因为他没有办法操纵水源污染的程度。但是他幸运地找到了一个自然情境，可以允许他在其他变量（主要是与社会经济水平有关的变量）受到控制的情况下，观察水源污染的变化。

让我们就拿斯诺的两组受试再来思考一下：一组由兰姆博斯公司供水，一组由南沃克—沃克斯霍尔公司供水。由于处在同一个地区，可能保证了两组被试的社会地位水平几乎相同。然而，斯诺所用的这一类型的研究设计的缺点是：它是由受试自己决定自己是属于哪一个组的。因为他们早在几年前已与两家公司签订了供水合同。我们因此还必须考虑的是什么原因促使他们与该家公司签约，而不是另外一家。是由于这家的价钱比较便宜，还是广告宣称他们的水质好、具有药

效？这些我们并不知道。关键问题是，选择那一家宣传水质有疗效的公司来供水的家庭，会不会在与健康相关的其他因素上也与其他家庭不同呢，而或许这些因素才是低发病率的真正原因。这是一种可能性。

斯诺的这一类型的设计也不能排除比较隐晦的虚假相关，这些虚假相关可不像前面所述的那些与社会经济因素挂钩的那样容易辨认。这也是为什么科学家情愿直接操纵他们感兴趣变量的原因。当操纵与一个叫做随机分组的程序（在这一程序中，受试将不能决定自己加入哪一条件的实验组，而是被随机地分派到其中一个组）相结合时，科学家就可以排除那些可以归于被试本身不同特征的解释了。当然，并不是说斯诺的方法没有价值，只是科学家会更喜欢直接操纵实验变量，因为这样推论才更具说服力。随机分组可以确保实验条件中的被试在所有其他变量上都相当一致，随着样本人数的增加，它还可以抹平掉一些偶然因素。这是因为这种分派方式是完全由没有偏私的随机抽样工具（例如，一枚硬币、随机数字表，等等）来选取样本，没有经过人手的选择。

至于古德伯格，他的确操纵了他研究假设中的原因变量。在他的一连串研究中，他不仅发现并记录了与糙皮病有关的变量，还直接控制了研究中的其他两个变量。请回想一下，他安排了低蛋白饮食的囚犯组来诱发糙皮病，以及吞食糙皮病患者排泄物的志愿组，其中包括他和他的妻子。这样，古德伯格不仅观察了自然存在的相关，还创设了特殊条件组，从而可以排除更多不同的可能解释，也因此他的推论就比斯诺的更有说服力。这就是科学家为什么试图操纵变量并保持其他变量

恒定的原因：为了排除可能替代自己的解释的所有其他解释。

医学中有很多由于对实验缺少控制而导致错误结论的例子。罗斯（Ross）和尼斯贝特（Nisbett）提到一个发生在19世纪60年代中期的案例（Ross & Nisbett，1991）。门腔静脉分流术（portacaval shunt），曾一度是治疗肝硬化的常用方案。当1966年人们开始对其进行研究时，发现了一个有趣的现象。在没有控制组的研究中，有96.9%的研究都发现，医生们断定的这种治疗具有中等以上程度的有效性；在有控制组但没有随机分组的研究中（这并不符合真实验设计的要求），有86.7%的研究得出同样的结论；而在有控制组而又有随机分组的研究中，只有25%的研究得到同样的结论。可见，这项治疗的效果——现在我们已经知道它是无效的，被缺乏控制的研究数据极夸张地高估了。罗斯和尼斯贝特也指出，“非经标准实验程序所获得的结果是一种‘安慰剂效应’或者是因没有随机分组而产生的偏差”（p.207）。他们还讨论了在无随机分组的情况下，选择偏差（见第五章的讨论）如何能够产生虚假相关这一问题。比如说，如果选择治疗的那些病人具有“典型样本”的特点，或者他们的家属态度积极、更为关心他们的病情，那么，就算结果发现治疗组与控制组有差异，也不能说明是治疗本身有疗效。

作结论前有必要先获取一些“比较信息”，这种思维习惯似乎不是与生俱来的，这也就是为什么所有科学训练中都包含强调设计控制组的重要性的研究方法课程。然而，控制组的具“非鲜明性”——除了改变某个关键因素外，控制组与实验组完全接受同样的待遇，使得我们从表面上很难看出它有

多么重要。心理学家曾经做过不少研究，来查看人们为什么这样不重视比较（控制组）的信息。例如，在很多研究中（Shanks，1995；Stanovich & West，1998），我们给被试呈现一个 2 × 2 的实验数据矩阵：

	有疗效	没有疗效
治疗组	200	75
无治疗组	50	15

表中的数字代表每种情况的人数。具体地说，这个表告诉我们有 200 人在获得治疗的条件组中接受了治疗之后呈现病情好转的现象；75 人则没有呈现好转；50 人是在没有接受治疗的条件组却在后来也呈现病情好转的现象；另有 15 人既没接受治疗也没呈现好转。研究者让看过这一矩阵的被试，来判断一下这个治疗的有效程度。结果发现非常多的人认为这个治疗有效，甚至有相当多的人认为这个治疗非常有效。这是因为他们首先注意到的是人数最多（200 人）的那一组，它正是接受了治疗并见好转的那一组。接着，他们注意到获得治疗的“治疗组”，治后呈现好转的人数（200）比没有呈现好转的人数（75）要多得多。所以就下了结论，认为这一治疗是有效的。但是实际上，这一治疗方案根本就是没有效果的。

此话怎么说呢？为了弄清楚为什么这个治疗是无效的，我们必须重点地看一下代表没有接受治疗的控制组（没有获得治疗的人）的两格数据。我们可以看到，控制组的 65 人中有

50人，或者是76.9%的人，即使没有得到治疗却也呈现了病情好转。这与275人中有200人，或者是72.7%的人，在接受治疗后呈现好转相比，很显然，控制组的好转率还大于治疗组！这就表示，这种治疗总的来说是完全没有疗效的。可见，由于人们经常忽视无治疗控制组的结果，并且只将注意力放在治疗组数据中的最大值和好转率上，这使得许多人会认为这一治疗是有效的。总而言之，人们很容易忘记：控制组的结果是在我们对治疗组结果进行解释时，非常关键的一个背景信息（见Doherty，Chadwick，Garavan，Barr，& Mynatt，1996）。

聪明汉斯的故事

行为科学中有一个著名的故事，可以很好地说明利用实验控制来消除对现象的各种不同解释的必要性：这就是有关聪明汉斯的故事，它是一匹会算数的马。80多年前，德国的一名中学教师向大家介绍了这匹会做算术的马。当汉斯的训练师向它呈现加、减或乘法运算的题目时，他会用蹄子敲出令人惊奇的准确答案。

许多人被聪明汉斯的表现惊呆了，而且迷惑不解：难道它真的能做过去人们认为马类不能做的事吗？你可以想像当时人们是怎么样的惊讶。对汉斯的独特能力的强力见证在德国媒体被广泛报道。柏林的一家报社记者说："这匹会思考的马将会让科学界在对许多问题做长时间的思考"（Fernald，1984，p.30）。这一预言很快成为事实，尽管其结果并不是那位记者当初所预期的样子。一组"专家"对汉斯进行观察

并证明了它确有这一能力。因此所有人都更困惑了。这种现象只要不实施控制比较研究，而一直仅是通过对孤立事件的观察，困惑就将一直持续下去。幸亏心理学家奥斯卡 · 方格斯特（Oskar Pfungst）对汉斯的能力重新进行了系统研究，谜团方才得以解开（Spitz，1997）。

方格斯特继承了实验设计的优良传统，系统地操纵了动物表演的环境。这样就可以创造出一些“人工”环境（详见第七章），从而用之来验证可以解释汉斯之表现的各种说法。经过许多仔细的测试，方格斯特发现这匹马确实具有一种特殊能力，但并不是算数能力。事实上，这匹马更接近一位行为科学家，而不是一位数学家。这是怎么回事呢？原来，汉斯是一个非常用心审视人类行为的观察者。当它要敲出它的数学答案时，它会观察训练员或其他提问者的头。当汉斯快要接近答案时，训练员会不自觉地歪下头，这时汉斯就会停下来。方格斯特发现马对视觉所提供的线索非常敏感，它可以察觉非常细微的头部运动，于是方格斯特想出了另一个测试马之能力的方法：他让不知道问题答案的人呈现问题，或者让训练员在马的视野以外呈现问题，结果，汉斯便失去了它的“算术能力”。

其实汉斯的故事是一个很好的例子，它向我们揭示了慎重区分对现象的“描述”和对现象的“解释”之重要性。马能够敲出训练员提出的算数题之答案是不争的事实，训练员也没有撒谎，而且许多旁观者也见证了这匹马确实准确地敲出了训练员所提出算题的答案。问题出在下一步：由该马能敲出正确答案推论出其具有算术能力！的确，下结论说马具有算术

能力是对现象的一个“假设解释”，而且它是众多可能解释这一现象的说法之一。因此从该马能正确地敲出算数题答案就很快地推论说其具有算术能力是言之过早。这一解释必须先付诸实征验证才行。一当它被付诸实征验证后，就被证实是完全没有根据的了。

就在方格斯特进行他的研究之前，对汉斯进行观察的专家就已经犯了一个根本的错误：他们没有看到对于马的神奇表现可以有许多不同的解释。他们认为，只要训练员没有撒谎，汉斯又的确敲出了算数题的正确答案，就足以说明马具有算术能力。方格斯特对此却进行了更科学的思考，他意识到这一解释仅是对该马的行为所可以做出的众多可能解释之一，而有必要设计受控制的实验情境来区分各种不同的解释：该马有算术能力，或是该马有察觉能力。于是方格斯特设计了一个让训练员在隔板后面向马提问这样一个情境来区分两种可能的解释：该马具有算术能力，或者该马是依据视觉线索来作出反应。如果汉斯确实具有这种神奇的算数能力，那么让训练员在隔板后面来提问，应该对该马的表现不产生任何影响，反之，如果该马只是依赖视觉线索来做反应，那么训练员在隔板后提问就会影响该马的表现了。一旦后者出现，方格斯特就可以排除该马具有算数能力的假设（Spitz，1997）。

这里可以和第三章讲到的“节约原则”联系起来——这一原则指出当两个理论具有相同的解释能力时，较简单的（涉及较少的概念及概念之间的关联）那个理论是比较可取的。这里存在争论的两个理论——一个认为汉斯真的具有算数能力，另一个认为该马在辨别行为线索——在节约原则上是截然不同。

后者不需要对先前的心理学及大脑理论做出巨大的调整，仅仅要求我们对马在视觉线索上可能存在的敏感性（现在看来这一解释的可能性很高）之看法稍加调整即可。而前一理论——认为马真的能运算——则需要我们对进化论科学、认知科学和脑科学等方面的十来个概念进行较大的修改。这一理论显然是极端的不节约经济。因为，如果我们认为这一理论是正确的话，那么，许多其他科学的概念也要跟着做彻底的改头换面才行（这就是将在第八章讨论的“关联原则”）。

20 世纪 90 年代的聪明汉斯：一个悲剧的衍生

聪明汉斯的故事经常在方法课上用来说明必须采取实验控制这一原则的重要性。不幸的是，人们不仅有时忽视了聪明汉斯个案的重要性，甚至那些给我们直接教训的个案看来也被忘得一干二净了。在整个 20 世纪 90 年代早期，全世界的研究者都在惊恐中观望——就像观看汽车被撞毁的慢动作——眼看着聪明汉斯的一个现代版在他们的眼前发生，并且导致悲剧性的结局（Gardner，2001；Spitz，1997；Twachtman-Cullen，1997）。

自闭症是一种严重的成长时期的障碍，其典型特点有：人际交往有缺陷，严重语言发展滞后及异常，活动及兴趣的范围狭窄等等（Frith，1989、1993）。由于许多极度缺乏人际沟通能力的自闭症儿童从外表看起来都很正常，以致使他们的父母特别不能接受他们的子女是有病的这一事实。这就不难想像，为什么在 20 世纪 80 年代末 90 年代初，当患自闭症儿童的父母听说澳大利亚有人发明了一种可以使过去从不用言

语沟通的自闭症儿童与人进行言语交谈的技术时，他们兴奋了。这种可以使缺乏口语能力的自闭症患者重获沟通能力的技术叫“受辅助的沟通”（facilitated communication）。当时，有相当一些极具影响力的媒体节目，像《60分钟》、《游行》杂志及《华盛顿邮报》等，都对此进行了毫无批评性的宣传报道（见Dillon，1993；Jacobson，Mulick & Schwartz，1995、1996；Mulick，Jacobson & Kobe，1993；Spitz，1997；Twachtman-Cullen，1997）。据这个技术的发明者称，自闭症患者以及其他因成长性能力缺陷导致不能讲话的儿童，只要把手和胳膊放在这台“具有同情心”的“辅助器”上，就可以在它的帮助下，在键盘上敲出相当有文采的句子来。可以想像，当部分先前言语能力极差的自闭症儿童做出这样惊人的言语表现时，他们深受挫折的父母会受到多大的鼓舞。这个发明家还声称，这个技术还可以帮助有严重智力障碍的失语儿童。

尽管作为父母的这种兴奋我们很容易理解，但是专业人员也这么轻易地就接受了这个发明却真的非常令人费解。很不幸的是，在尚未开展任何控制和研究该项技术的疗效之前，其功效已经通过大量的媒体传播给了许许多多满怀希望的父母们。如果有关的专业人员在实验控制上稍稍具有最基本的训练，就应该马上意识到它与聪明汉斯个案之间的相似性。这种辅助器可以说是一个永远关注儿童成功的、富于同情心的“人”，在辅助过程中有许多机会，在意识层次或下意识的层次，引导儿童敲击键盘。其实，受到辅助器所给线索的引导这一解释也被另一个观察到的现象所支持。儿童有时在不看

键盘的情况下，就可以打出复杂难懂的句子。另外一个支持这解释的观察是，一些从没有见过字母的幼童也可以用这个辅助器创作出具有更高难度的英文散文。

一些有控制的研究报告称，他们通过适当的控制对受辅助的沟通技术进行了验证。每一个研究都清楚地得到了同一结论：自闭症儿童的表现依赖于辅助器发出的、不易被察觉的线索（Beck & Pirovano，1996；Burgess，Kirsch，Shane，Niederauer，Graham，& Bacon，1998；Cummins & Prior 1992；Hudson，Melita & Arnold，1993；Jacobson et al., 1995，1996；Levine，Shane & Wharton，1994；Prior & Cummins，1992；Shane，1993；Simpson & Myles，1995；Smith & Belcher，1993；Spitz，1997；Wheeler，Jacobson，Paglieri，& Schwartz，1993）。这些研究中，有一些所运用的控制都与聪明汉斯这个典型例子中所用的相似。他们所设计的控制情境是给儿童和辅助器各自呈现一个画出来的东西，但是他们彼此却看不到呈现给对方的东西是什么。当向儿童和辅助器呈现同一个东西时，儿童打出的是自己看到的东西的名字。但是当给儿童和辅助器呈现不同的东西时，儿童打出的却是呈现给辅助器看的东西的名字，而不是他自己正在看的东西。也就是说，儿童的回答取决于辅助器，而不是孩子自己的能力。得知“受辅助的沟通技术”只不过是一个“聪明汉斯”现象，而不是突破性治疗技术的结果这一结论，并没有使这些研究者为之雀跃。因为这一“聪明汉斯”现象远远不只限于在这一辅助器之疗效研究中发生，许多其他研究都有同样的问题。心理学家卡明斯和普赖尔（Cummins & Prior，1992）结论说：

"很明显的，许多研究助理都是在运用了不易察觉的视觉线索，或是运用自身的动作姿态等，来控制他们病人的反应，从而得到有疗效的数据。这是目前研究证据所给我们的不想要的，但却又是不可避免的结论"（p.240）。

这一不幸的事件愈演愈烈。在一些治疗中心，有当事人在接受辅助器帮助的沟通过程中，讲出他们曾在过去遭到父亲或母亲性虐待的经验（Dillon，1993；Seligmann & Chideya，1992；Spitz，1997；Twachtman-Cullen，1997）。于是这些孩子们被迫从自己的家中搬出来，直到这种指控被证明是无中生有之后，才再搬回去住："许多法院的诉讼案件，皆因受辅助的沟通技术所引起，其中最引人注意的是几起与虐待有关的案件。当法院下令进行的察证都证实是辅助器严重地影响了人们的沟通之后，这些指控就都被取消了"（Smith & Belcher，1993，p.176）。赫德森（Hudson）等人（1993）报告了对一名妇女的测验。这位29岁妇女被认定是极度智力低下，一直和父母生活在一起。在接受"受辅助的沟通"治疗期间，她报告说曾遭到生活中关键人物的性侵犯。在进入法律程序时，赫德森等人运用了先前研究者，使用了聪明汉斯研究中所用的逻辑，重新测试了该妇女在接受辅助器治疗时的沟通能力。通过两个不同的耳机，向她和她的辅助器读出问题。当两者被问到的问题相同时，每次她都能够正确的回答问题；但是当两者被问的问题不同时，她从没有给过正确的答案，而且其中40%回答了辅助器被问的问题。

就像《新闻周刊》报道的：

这种未经证实的控诉已经造成了很大的伤害。父母不得不

在忍受邻居的责难和雇佣律师的高额费用下，让人把自己的孩子带走，安置在临时监护人的家中——而最后换来的只不过是学校几个月后寄来的一封道歉信而已。被牵连的老师和他们的助手也因为这一丑闻的阴影，而断送了他们的事业。（Seligmann & Chideya，1992，p.75）

当越来越多的证据发现这种缺乏实征基础的治疗并不只是无害而已时，人们终于在媒体的喧扰之下，开始注意到了专家的意见。（“噢，也许，它是有效的，但是假如它是无效的，那怎么办？”）。推广这种尚未证实的治疗技术所带来的社会成本却是千真万确的。波士顿儿童医院的沟通促进中心主任夏恩（Shane，1993）曾说：

所有的科学迹象表明，受辅助的沟通技术是无效的。……由于这种方法所可能带来的潜在危害，所以我们很难建议继续使用它。例如，在接受辅助器治疗期间发生的性虐待指控，教学人力资源安排的不当，以及为了要启用这项无效技术而拨出的巨额培训和研究经费（p. 11）。

俄亥俄州立大学儿科及心理学教授缪利克（见 Mulick，Jacobson，& Kobe，1993）更是一针见血地指出了因这一治疗“一窝风”所付出的代价：

如果没有对 FC（受辅助的沟通技术）的大力宣传，我们可能可以得到更多的人力和金钱用于发展基于更有实征基础的、更可行的长远策略来解决困扰儿童的这一问题。FC 的拥护者为研究和专业文献所带来的理论上的混乱，损害了对能力

> 缺陷及其成因方面知识之累积，也引开了那些真正有心有力想要将成长障碍的研究加以综合的人之注意力。由FC而引起的混乱，甚至使其他已经成功运用于残疾人群的非言语沟通系统也普遍遭到公众的排斥……我们的经验告诉我们，残疾人士可以成为其家庭和社区中有价值的一分子，而无需去求助于神医神药。他们可以寻求现有的、有科学依据的有效帮助。那些具有科学训练，富有热情的专业人士的努力，绝对也永远要比各种瞬间即失的治疗热卖要好得多。治疗的进步及知识的进展是建筑在严格的训练、对正确的及科学标准的尊重，及对治疗疗效的客观论证之上的（pp.278～279）。

以上这个例子再次说明了，因迷信见证叙述所可能造成的伤害，以及认为治疗热卖与伪科学是无害的，这一观点所可能带来的谬误（见第四章）。由此，我们也可以看到，要想正确地解释行为，实验方法中的控制和操纵是必要的，也是不可替代的。

这里需要再次强调的是“节约原则”与控制和操纵的关联。如果患有严重言语障碍的自闭症儿童可能通过一个“魔弹”（见第九章）的干预而得以治愈，那么它显然可以推翻过去几十年来对患自闭症儿童所做的认知、神经心理学以及大脑特征等领域的所有研究成果（Baron-Cohen，1995；Frith，1993；Leslie，2000；Sigman & Ruskin，1990；Yirmiya，Erel，Shaked，& Solomonica-Levi，1998），从而我们必须要对已取得的认知和生物知识做出很大的修正。受辅助的沟通技术与其他的科学研究成果几乎是完全没有任何关联性（见第八章）。

用精心设计的情境来区分变量

古德伯格的例子给我们上了重要的一课，有助于我们消除科学进程中的许多错误观念，特别是在心理学的研究上。世界上所发生的许多事情经常都与很多其他因素相关，为了仔细区分和辨别出许多同时发生的事件之因果影响，我们必须设计出一些一般在现实世界中从来没有的情境。科学实验就是要打破自然界原有的各种相关性，而将单一的变量所可能造成的影响孤立出来。

心理学家们不例外地，也采用了同样方法：通过操纵和控制来分离变量。例如，对阅读过程感兴趣的认知心理学家，已经研究了什么因素使得人们对文字的感知更容易或者更困难。毫不奇怪地，他们发现长的单词比短的单词更难确认。乍一看，我们可能认为字的长度应该很容易测量：只要收集两组单词，一组长的，一组短的，然后测量被试在阅读这两组词时的速度差异就可以了。但说起来容易做起来难。长字在语言中出现的频率小，而频率本身就会影响我们对词的感知。因此，长短词中出现的任何不同既可以归因于长度，也可以归因于所用单词的频率，或者是两者的交互作用。为了看一下在不受频率影响的情况下，字长是否影响感知，研究者就必须构建特殊的单词组，以确保字长和频率不会同时发生变化，以致同时影响对字的感知。

与此相似，古德伯格之所以可以对原因做出有力的推理，是因为他设计了一系列在自然环境中不会发生的特殊情境（回想一下，他的一个实验操纵是要被试吃下人体排泄物！这是何等的

不自然呀！）。回想方格斯特同样地为测试聪明汉斯的能力而设计的一些特殊情境，包括提问者不知道问题答案的情境。许多人只观察到马在正常情境下回答问题（亦即在提问者知道答案的情况下），但是从来没有去想，马是怎么会做出这样惊人之表现的，反而做出了马真的具有算数能力的错误判断。

同样，在检验受辅助的沟通技术的功效时，研究者也必须设计一些特殊的情境。展示给辅助器和儿童面前的刺激物必须分开来呈现，并且不让任何一方知道对方的刺激物是什么。这种特殊情境对检验解释同一现象的不同假设有着至关重要的意义。

许多心理学的经典实验都包含了这种分离现实中之自然联系的逻辑，来判断哪一变量是现象的决定因素。心理学家哈里·哈洛（Harry Harlow）的著名实验（Anderson & Anderson，1996；Harlow，1958）提供了这样一个相关的事例。当时，关于母婴依恋这一现象有一种很流行的假设：这种依恋是来自母亲是婴儿食物的来源。然而，问题是，母亲能提供的当然不只是食物而已（还有舒适、温暖、爱抚、刺激等）。也有可能是这些其他的原因造成了这一依恋。哈洛（1958）正是想对这一问题进行验证。他选用了猴子来进行他的研究，并设计了一些"人工"的情境，在每一个情境里，除了一个可能造成母婴依恋的原因在作用之外，其他变量都受到了控制，让它们在所有情境保持不变。他让在这近乎隔离环境中的婴猴选择与一个"人造"母猴在一起生活，并观察它们的行为。他发现，由绒线做成的人造"母猴"给予的触摸舒适远胜于用铁丝网做成的"母猴"。在婴猴两个

星期大的时候，它更愿意接触冰冷的绒线母猴，而不是温暖的电线网做成的母猴。这个发现表明，由触摸所带来的舒适感远比温暖感更重要（Harlow & Suomi，1970）。最后，哈洛还发现，即使当婴猴的食物完全由电线网做的母猴提供，它仍选择绒线母猴。因此可以证明，依恋仅归因于母亲能提供食物的假说是错误的。哈洛能做出这些结论是因为他可以把现实世界中同时变化的变量分离开来。

总之，对于科学家来说，检验解释某一现象的特殊假设或理论时，设计一些特殊的情境是必要的，仅仅在现实世界的自然状态下观察事物是远远不够的。人们观察下降和移动的物体已经有几个世纪了，却没有人得出移动和重力的正确规则和定律。直到伽利略和其他科学家设计了一些人为情境来观察物体运动时，才得到了真正能解释运动的定律。在伽利略的时代，很少有人会去看光滑的铜球从光滑的斜板上滑下。尽管在这个世界里有许多运动在发生，但很少有以这种特殊形式出现的。然而，正是像这样的一个非自然的情境，加上其他与此类似的情境，促使我们发现了关于运动和重力的第一个真正解释性的定律。讲到运动定律，你不是在这一章的开始，做了一个小小的测验吗？

直觉物理学

其实，本章开头提到的3个问题来自于约翰·霍普金斯大学的心理学家米歇尔·麦克科劳斯基（Michael McCloskey）的一本著作。麦克科劳斯基（1983）研究过他所称的“直觉物理学”，亦即一般老百姓对于物体运动的认

识。有趣的是，这些认识经常与物体的真实运动完全相反（Catrambone，Jones，Jonides & Seifert，1995；diSessa，1996）。

比如说，在第一个问题中，一旦系住正在转圈的球之绳子被切断，球会沿与绳子成90度（圆的切线）角的直线飞出去。麦克科劳斯基发现，1/3的大学生错误地认为球会沿抛物线飞出去（也见Catrambone et al.，1995）。对于轰炸机飞行员的问题，大约一半的人认为炸弹应该在目标的垂直上空抛出，这足以看出人们对于物体初始运动决定其后来运动轨迹这一知识有多缺乏。实际上炸弹应该在飞机距离目标还有5英里时抛出。而被试的错误并不是由于这些题目都是一些假想的抽象情况，太难让人回答了。因为，当要求他们在屋子里边走边丢一个高尔夫球时，几乎一半以上的人根本不知道球在下落时会向前移动。至于最后一道题的答案，许多人都没有想到，子弹从枪管中发射出来，与子弹从同一高度下落会同时到达地面。

你现在可以算一下你这次回答本章开始时所提出的问题之成绩。如果你最近没有上物理学课的话，你有可能至少错一题。"物理学课！"你可能会抗议，"我最近当然没有学过物理学，当然答不对了！这个测验太不公平了！"但是，喂，你为什么需要上物理学课才知道这些题目的答案呢？在你的生活中，看见物体从高处掉下恐怕不下几百次了吧？而且它们都是在自然的情境下掉落下来的。每天在生活中，你也见过不少物体在运动，而且你是在它们的"真实"状态下看到它们在运动的。所以，你当然不能说你从来未看到过正在运

动的，或是正在下落的物体。就算你辩称从未看到过像子弹飞出这样的事件，但是我们几乎都看到过孩子们将旋转的物体放开的情境，而且我们中很多人也都见过从飞机上扔东西的情境。你说你从未见过完全类似的情境，似乎是有点强词夺理吧？既然你每天都会遇到这些运动或下落的物体，为什么你还是不能准确的预测那些与常情稍有不同的情境会发生什么事情呢？

麦克科劳斯基的工作向我们表明了理解科学家为什么还要做他们做的那些事，这是非常重要的。不管人们对于运动和下落物体的经历有多么多，他们对于运动的直觉理论仍非常不准确。关键是我们要理解外行人之所以会有那些错误的观念，正是因为他或她的观察是在"自然的"情境中得到的，而不是用科学家的方式在控制的情境中进行的。因此，如果你在本章开头的小测验中回答错了，也不要感觉自己是无知的，或者不够好的。要知道，一些世界上伟大的思想家都已观察物体下落好几个世纪了，但是对运动物理都没能想出什么比高二的学生所能想出来的更准确的理论。在《美国科学人》杂志上的一篇文章中，麦克科劳斯基（1983）观察到，他的很多被试对运动的物体所持有的不正确看法，竟与牛顿前3个世纪的想法不谋而合。麦克科劳斯基的现代被试和中古世纪的哲学家在某些方面是相同的：两组人都是在自然世界中对物体运动做观察，而没有运用人为设计的情境来进行科学的操纵、控制和比较。

即便是大量的个人经历，也不能防止对物体运动本质是什么的错误观念的产生。飞行员兰格威斯基（Langewiesche，

1993）在阐述关于飞机转弯角度这一知识的发展史时指出，20世纪早期的飞行员对于使用旋转器这类装置非常排斥，因为他们相信"平衡本能"。然而，在多云的天气里，对陆地的能见度低，因为不能借助"本能"飞行员一直感觉不到岸边，因而无法正确地将飞机旋转到想要的角度。许多飞机坠毁及险些坠毁的事件之后，飞行员终于领悟了一个沉重的教训：再多的直觉，也不能取代飞行物理学的知识（Langewiesche，1993）。

直觉心理学

哲学家丘奇兰德（Churchland，1988）论及，如果我们关于运动物体的通俗理解都是不正确的，那么很难相信我们对有关人类行为这一更复杂的领域的通俗见解会是正确的：

我们最初关于运动的通俗理解是非常模糊及混乱的，而且最终将被更成熟的理论完全取代。我们早期关于理论的结构及活动的通俗理解也是完全离谱的，它们之所以依然存留下来，只不过作为一些历史的教训，让我们知道自己可以错到什么地步而已。我们关于火的本质以及生活本质的通俗理解也同样愚蠢。由于我们过去的许许多多的通俗想法都被推翻了，所以我们可以这样一直数落下去……但是有关人类的心智活动这一现象要比刚才提到的任何一个现象都要复杂和难以理解。如果我们对其他事情的理解都做得这么糟糕，那么，想要在一开始就能正确地认识心理学知识，简直就是天方夜谭（p.46）。

生物学家E · O · 威尔逊（E. O. Wilson） 在赞同丘奇

兰德的推测的同时，指出“大脑这部机器，不是为了理解它自己的，而是为了延续生命。这两个目的在根本上完全不同，如果没有从科学中所获得的事实知识的帮助，大脑也只能看到世界的一些小碎片。为了要考虑明天如何能继续生存，它必须把注意力放在它必须知道的世界的一角……这就是为什么即使到了今天，人们了解他们的汽车比了解他们的大脑还要多——以及为什么对于人类思维的基本理解还是停留在是一种实征性的层次，而并没有上升到哲学的或宗教的层次”（Wilson，1998，pp.96～97）。

当我们审视有关人们行为理论的现代文献时，我们可以发现丘奇兰德的推测是正确的。在第一章中，举例说明了一些对人类行为的普通常识性（或通俗性）理解是错误的。在奥费·科恩（Alfie Kohn）的著作《你知道他们在说些什么……：通俗信念的真相》（1990）和色吉奥·达拉·萨拉（Sergio Della Sala）的著作《心灵神话：对思维及脑之通俗假设的探研》（1999）中有更多这方面的例子。尽管两本书都是为普通老百姓而写的，他们还是大量地采用了在有同行进行审查的科学学术杂志上发表的论文，作为他们结论的证据。在这两本书中讨论的许多通俗理论都与人类行为有关，例如，没有证据表明宗教信仰的虔诚度比较高的人比虔诚度比较低的人更有做利他行为的倾向（Paloutzian，1983；Smith，Wheeler，& Diener，1975）。许多研究已经表明：宗教信仰的虔诚度与其参与慈善活动、帮助处于困境中的人，或者不欺骗他人等行为倾向之间没有简单的关系（Paloutzian，1983）。相反地，在大量研究文献中，也没有迹象表明高宗

教虔诚度的人比那些把自己标榜为无神论者的人更有善心，或者更愿去帮助别人。

有关通俗信念不正确的例子可以举出一大堆：例如，许多人认为“月亮的盈亏影响人的行为”，其实这是不对的（见Byrnes & Kelly，1992；Coates，Jehle，& Cottington，1989；Culver，Rotton，& Kelly，1988；Rotton & Kelly，1985）。有些人认为“异性相吸”，他们也错了（见Buss，1985；Buss & Barnes，1986；Murstein，1980）。有些人认为“熟悉产生鄙视”，这也是错的（见Bornstein，1989）。还有人认为盲人天生赋有超常敏感的听觉，情况并不是如此（见Niemeyer & Starlinger，1981；Stankov & Spilsbury，1978）。这类的例子举不胜举。有兴趣的读者可参考科恩（Kohn，1990）的书，他在书中详细列举了大量不被实征证据所支持的有关人类行为的通俗信念（当然，这里列举的所有结论都是建筑在统计学“概率”的基础上的，个别案例可能与这些结论正相反：见第十章）。

人们有关行为的直觉理论的许多不足之处，说明了为什么我们需要有控制的心理学实验：这样我们才能走出对人类行为的古老想法，走向建立更为准确的科学构念。

小　结

实验法的中心思想是操纵及控制。在相关研究中，研究者仅仅观察两个自然变量间的变化是否存在联系；而在真正的实验中，研究者操纵那个假设可能是一个现象变量的原因之变

量，并通过控制和随机分组来维持所有其他可能与该原因变量相关的其他变量的恒定，以寻找原因变量对现象变量的真正影响。这种方法排除了相关研究中出现的第三变量问题。第三变量问题的出现是因为在自然世界中，许多不同事物是相互关联的。实验方法可以看作是用来区分这些自然联系的一种方式。之所以这样做是因为通过操纵假定的原因变量和恒定所有其他变量，这样可以将假定的原因变量分离出来单独审视其对现象变量的影响。但是，为了要分离这些自然联系，科学家们经常被迫设计一些在自然世界中找不到的特殊情境。

（马淑蕾　译）

第七章 “但，真实生活不是这样的啊！”

——“人为性”的批评与心理学

前两章中，阐述了心理实验的基本逻辑思路。现在再了解一下经常听到的，针对心理学的种种批评。通过对这些批评的讨论，我们可以加深对科学方法的理解，而本章特别需要深入探讨的是，由心理实验是“人为的”或“不是真实生活的”，而否认实验方法是有用的之类的批评。理解这类批评不能成立的原因，将有助于我们更好地了解心理实验，进而了解心理学。

自然性有时不是必需的

经过第六章的讨论，已经很清楚这一点了吧。正如前面提到的，科学实验的人为性不但不是什么毛病，反而，正是这人为性才使科学方法拥有解释自然世界的巨大力量。与人们的一般认识不同，科学实验的人为性是科学家们深思熟虑后的结果，而非一时疏漏。科学家精心地设计一个非自然的环境，以便将那些众多可能决定事件发生的相关变量分离开来。

有时，起决定作用的条件在自然状态中已经存在，如斯诺与霍乱的例子，但这种情况毕竟只是少数。而更多的时

候，就需要科学家通过一些自己设计的，有时甚至是新奇的方法，才能操控事件的发生，如古德伯格与糙皮病的例子。在许多情况下，这些操控手段是不会在自然环境下发生的，因此科学家就必须将其引入到实验室中来，从而可以进行更精密的控制。例如，在对“重力与运动”的早期研究中，为了制造出某种特殊情境以便观察物体运动，就构建了许多物体。这就说明，为了分析某种现象，在实验研究中设计一个“非自然”甚至非常“极端”的情境是很有必要的。

实际上，如果科学家只是严格地局限于自然观察，那么许多现象就完全没有可能被发现。例如物理学家为了要研究物质的根本特性，造出了一个长达1000米、巨大无比的加速器，用来诱发微粒子之间的碰撞，在这些碰撞产生的过程中，他们发现了一些未知的、存在时间不到十亿分之一秒的粒子。而这些新粒子的特性对原子结构的理论有很重大的意义。但是这些新粒子多数在自然情境下根本就不存在。即使存在，也不可能用自然观察法察觉得到。很少会有人对物理学家的这种做法提出怀疑，用特殊的、有时甚至是稀奇古怪的方法来探究自然，被看成是深入了解宇宙的有效途径。然而，当这种对物理学家而言很有道理的做法一旦应用到心理学上，就被看成是无效的了。

当人们不了解在实验室中设计特殊情境的重要性时，往往会误解科学家们的做法，而心理学家更是处在误解的重灾区。当心理学家拿出某一行为在实验室中所收集的证据时，经常会听到外行人这样的悲叹：“但，真实生活不是这样的啊！”。当我们深入探讨这句话时，会发现这些人坚持认为

只有在自然情境下做的研究才是真正的研究，而心理学的实验研究是怪异的，因此也意味着心理学不足以作为一门科学。

人们一般不会意识到，心理学家所采用的那些看来有些奇怪的手段，也并不单只是心理学才用，其实这些手段只是把科学方法应用来研究人类行为而已（Banaji & Crowder, 1989）。这些为了了解世界而使用的奇特方法，同样会在任何一门科学中看到。但不幸的是，心理学家却因此受到了双面夹击！除了这些人控诉心理学家不用自然情境来研究行为，因此是不能反映真实的生活之外，还有另外一些人，同样因为对研究方法的无知而持有认为心理学永远不能成为科学的信念，因此藐视心理学家，虽然心理学家们也像其他科学的研究者一样，在通过设计必要的、特殊的情境来发掘对某一现象更加有力、精确的解释，他们却视而不顾。

对科学研究要求只局限于观察真实生活，会妨碍我们发现许多事情。例如，现在已经广泛应用于众多领域，如周期性偏头痛、紧张性头痛的控制，高血压的治疗以及放松训练等的生物反馈技术（Miller, 1985a），一开始就源于一个实验研究的结果：即当人们能够得到视觉或听觉的反馈，从而用来监控自己内部的生理功能运作时，就能部分地学会控制这些过程。当然，由于人类本身没有这种通过外部反馈来监控内部生理功能的器官。所以这种控制能力，也就只能在特殊的实验情境下才会表现出来。若只通过自然观察来研究这个问题，就永远也不会发现这一现象了（Henshel, 1980）。正如，安德森、林赛（Lindsay）和布什曼（Bushman）所说的："大多数实验研究的目的都是为了要发现，一些构念变

量之间有哪些理论上的关系，而这些变量又是在真实世界中不能被充分地分离开来以便作精确审察的”（Anderson，Lindsay & Bushman，1999，p.4）。

对“随机取样”的误解

有时，人们对“但，真实生活不是这样的啊！”的抱怨是缘于对心理实验目的的误解，其实它并不难理解。媒体的宣传使大家对调查研究都很熟悉，尤其是对正式选举中的民意调查。并且也渐渐意识到了选举调查中的一些重要特性，特别是媒体给予格外关注的随机取样与样本代表性问题。人们知道这两个特性对于民意测验的准确性起着关键性的作用。这也就使许多人错误地认为随机取样及代表性是所有心理学研究的必要条件。由于心理学研究中很少使用随机取样来选择被试，所以那些认为只有使用随机取样才行的外行人，就开始指手划脚认为心理学研究由于没有用随机取样的样本，因而没有反应真实生活，也因而是没有用的。

但只要想一下其他科学的情况，就很容易理解这种想法的荒谬。化学家从未试过要将化合物随机取样；生物学家也不会做什么关于细胞与器官的随机取样；在医学研究中应用的老鼠和猴子等也不见得就能代表他们的种族。而这些研究都是在与这些动物生活的自然环境完全不同的实验室中进行的。实际上，这些研究情境可以说绝对是独特的，然而所有这些研究所得到的结果，却可以帮助理解生物学。多数心理学的研究也是这样的，并不是每一个心理学的研究都必须用随机取样来找研究对象。

理论导向的研究与应用研究的异同

那么为什么对于选举那么重要的随机取样，对心理学研究而言就不那么重要了呢？对于这个问题，可以通过考察各种应用研究与大部分基础研究之间的不同，来寻找答案。维吉尼亚大学的心理学家道格拉斯·默克（Douglas Mook），根据不同研究类型所要求的不同预测层次对上述问题进行了阐述（Mook，1983，1989，2001）。许多类型的应用研究的目的就是要将研究的结果直接应用到日常生活中的一个特定情境。对应用研究而言，预测必须是：研究与现实生活有“一对一”，亦即默克称作“类比”的关系。也就是说，研究结果必须能直接应用才行。选举中的民意测验便是一个应用研究来做“类比”预测的例子，其目标就是为了预测选举那天的投票结果。由于研究结果要直接被应用，样本的随机取样性及情境的代表性就显得格外重要了。

尽管应用心理学研究已经变得越来越普遍了，但是，如果由此认为这类研究是典型的心理学研究则是错误的。实际上大多数心理学研究（其实，在这一点上，其他学科也是如此）目的并不在于应用。其所关心的预测层次都是从理论到特定的研究情境，而且所得到的结果也不能直接应用，必须通过对理论的调整，再将理论和其他科学规律结合起来，才能应用到实际问题中。简而言之，大多数理论导向的心理学研究是为了要验证有关某一心理过程的理论，而不是把研究结果推广到现实中的某一特殊情境中去。

目的主要在于验证理论的研究，通常被称为是“基础研

究”。应用研究的目的是从研究数据直接走到现实世界中去，而基础研究则关注理论的验证。但是，如果仅仅根据一个研究是否能实际应用来分辨这两种研究的差别的话，那就可能太简单化了（见 Nickerson，1999，有关两者的相互依赖性）。因为这一差别常常是时间的差别所造成应用有早晚的问题。应用研究的发现可以立即被利用，然而没有任何东西比一个普遍的、精确的理论更有用，只不过这种理论需要时间的酝酿。在科学历史的长河中，有很多的例子让我们看到，尽管原本一个科学家并无意去解决什么实际问题，但是他的理论或发现最终还是解决了真实世界的很多问题。

在布鲁海文国家实验室这个处在领先地位的物理研究机构中工作的克瑞斯（Crease）和萨姆埃尔斯（Samios），曾在其文章中讲述了几个基础研究本身已经具有应用价值的例子（Crease & Samios，1991）。他们提到威尔赫姆·伦琴（Wilhelm Roentgen）的故事。伦琴“为了要弄明白为什么放在他仪器旁边的一个荧光屏幕会意外地自动发光，结果发现了自然界的一个新现象——他称之为X射线。3个月之后，X射线就被用来检查骨折了”（p.82）。再看一下霍华德·弗罗瑞（Howard Florey）的例子，他和一名同事从事抗菌机制的调研，“在他们要研究的微生物中有青霉素的名字，这种细菌在几年前才偶然被发现，但它的抗菌能力却一直没有被人注意到。由此可见，现代医学的进展很多都是建立在这类对应用没有特别兴趣的学术研究的发现和发展之上的”（p.82）。

西摩·凯特（Seymour Kety）也谈到了他是如何在看似毫无关联的科研中发现了治疗精神分裂症的一个特效药——氯

丙嗪（chlorpromazine）（Kety，1974）。凯特甚至说，实际上应用于精神分裂症治疗的任何一个发现，原本都不是为了治疗精神分裂症而做的研究！我们一次次地看到，试图去控制科学的研究方向（亦即，告诉科学家去解决一个实际问题的企图）往往造成正相反的结果：它阻碍了而不是促进了科学的进展。极具讽刺意味地是，积极想要科学只去解决实际问题，而不要用管“那些其他的事情”（基础研究）往往是最不切实际及最短视的做法。

通往实际应用的道路往往是充满着艰辛与意外的。得克萨斯西南大学的医学中心有一组研究人员为了研究关节炎，找来了一些患有关节炎的老鼠并进行了染色体干预。出乎意料的，他们的老鼠竟然也患了肠炎（Fackelman，1996），“我们没想到会将老鼠弄成溃疡性大肠炎，因而能为今后其他科学家研究人类疾病提供难得的有肠炎的动物样本”（Fackelman，1996，p.302）。不管他们在治疗关节炎上是否能有进展，但是他们看来却为溃疡性大肠炎的病因及治疗做出了巨大的贡献。

简而言之，我们必须认识到，尽管有一些研究是为了直接预测某一特殊情境而设计的，但是大多数的科学研究仍然意在验证理论的基础研究。回答“这些研究结果要怎么应用到现实生活中呢？”这一问题，从事应用研究和基础研究的工作者会有完全不同的答案。应用学者会说：“可以直接应用，只要实验情境与将来要应用的情境有相当程度的相似性就可以了。”因此，像样本的随机取样性及实验情境的代表性等问题都影响其结果的应用效果。然而，理论学者会回答，

他们的发现不能直接应用于现实生活，他们的研究目的也不在于此。因此他们这时也不关心他的研究被试与其他群体有多大的相似性，或实验情境是否反映了真实生活中的情境等这类问题。但是，这是不是说这些研究结果就不能应用于现实世界呢？当然不是。这些结果不能直接应用于现实世界，但是却能直接应用于一个理论。而一个理论则会在将来与其他科学规律结合之后，应用于解决一个现实问题。

这类通过理论而间接地应用于现实生活的例子在心理学的一些领域中比比皆是。例如，几年前移动电话刚刚进入市场时，不少认知心理学家就意识到如果人们开车时接听电话很可能会存在交通安全隐患，因此他们立刻预测使用移动电话将导致交通事故增多。他们推论的理由倒不是因为接听电话的时候手会离开方向盘，而是由于司机的注意力将转移到谈话上。值得注意的是，心理学家提出这些担忧远远早于真正用移动电话来做的实验研究（见 Strayer & Johnston，2001）。他们是用在心理学里已经存在有几十年历史的、关于注意力有限加工能力的理论做出这样的推测的。这一信息加工理论是经过了无数次实验验证并得到支持的（如，Broadbent，1958；Kahneman，1973），而开车时使用移动电话正好可以用上这一理论来预测其可能造成的危害。实际上，后来使用移动电话来做的研究也证实了心理学家关于注意力理论的假设：使用移动电话的确可能引发交通事故（Redelmeier & Tibshirani，1997、2001）。

默克（1983）提到两个例子可以用来说明，心理学中通过实验来验证理论的理念以及间接应用性的本质。20 世纪 30

年代，塞丽格·赫奇（Selig Hecht）在“普通实验心理手册”的杂志中发表了对视觉敏感度（视敏度）的一系列研究（Murchison，1934）。这些研究中谈到了适应黑暗现象。当我们走进黑暗的电影院时，会出现暂时的“失明”，但如果你在座位上待一会儿，你更会察觉到椅子及人，然后其他物体也都慢慢看得见了。如果你再仔细留意一下，你会发现这一视敏度不断升高的现象会持续几分钟之久。

这种现象被称为“暗适应”过程，它分为两个阶段：首先是刚刚进入一个黑房间时，视敏度迅速但小幅度的升高；在这之后，视敏度会大幅度的升高。赫奇将视敏度在这两个时段不同的升高曲线，与视网膜上的两种不同感光细胞联系起来。其中视锥细胞集中分布在中央凹（光线在视网膜汇聚的部位）的中心，它们对红光格外敏感。而视杆细胞位于中央凹的外围，分布不集中，对红光也不那么敏感。赫奇根据这些有关视网膜结构的知识，提出了一个理论，认为暗适应最初那个阶段（即视敏度迅速小幅度升高的阶段）是由视锥细胞的适应作用所引起，而第二阶段（即视敏度的大幅升高）则取决于视杆细胞的适应作用。

默克（1983）提醒我们注意赫奇的实验是在完全非自然的情境中进行的。被试（当然不是随机取样的）被关在一间暗室里，根据他们是否察觉到面前闪烁呈现的红光点，回答“看到了”或是“没有看到”这样的问题。这种微弱的红光点是我们在日常生活中根本不会理会的。然而，赫奇根本不关心被试在暗室中的“是”、“否”回答为实验情境是否能推广到一般老百姓的生活中，因此这种情境是否可以在现实

生活中找到，对他来说无关紧要。他真正感兴趣的是建立及验证有关视觉系统结构的理论。对像暗适应这样的视觉基本过程的研究，正可以帮助他建立支持他的理论的证据。他并不关心他的实验情境是否在现实中存在，而只关心这种情境能否可以有效地将他所要研究的这一特殊的视觉过程分离出来。

赫奇的研究发现之所以能取得普遍性，并不是因为他的实验情境的性质，是人工的或是自然的，而是因为他有能力建立一个有关视觉过程的理论，而这个理论可以与许多的视觉现象或运作相关联，以致他可以利用它们来验证他的理论。他的研究发现了人类视觉系统中各个部分之间的功能关系，正是因为他的实验情境是受控制的并且是人工的。如果他的功能关系理论是对的，那么它将来的应用价值应该是非常高的。并且可以用来解释与发现其他与这一理论的实验情境完全不同的情境中所出现的行为现象。

换句话说，赫奇的发现是通过它们对理论的影响而间接地找到了可以应用的地方。例如，通过赫奇所提出的视觉系统理论，已经对治疗夜盲症以及识别 X 射线产生了很大的帮助（Leibowitz，1996；Mook，1982）。更有意思的是，二战期间，那些等待德军夜间突袭的英国飞行员，都会戴上一副红色的飞行眼镜（因为视杆细胞对红色光线不敏感——从而在黑夜里可以停留在暗适应的状态，见 Mook，1982）。从让被试判断小红点，到伦敦上空危险的空战，这一鸿沟是通过理论来跨越的，而不用将赫奇的实验室重新设计成战斗机机舱的样子。

心理学理论的应用

一旦我们了解了大多数研究是为了发展理论，而并非是对特殊情境进行预测，并且也知道了大多数研究的发现是通过理论得到间接应用，并非是直接应用于特殊情境，那么我们可以问，心理学已经有多少通过理论得到应用的例子。也就是说，我们想知道，有多少心理学理论通过了普遍性的检验？

在这一点上，我们必须承认以往的记录是极其糟糕的，但也必须清楚地意识到这是与心理学的多样性息息相关的。一些领域的研究确实在应用方面的进展甚微，然而，其他一些领域则已经取得了十分瞩目的成绩，通过实验已经推导出了许多具有解释力及预测力的原理（见 Byrnes，2001；Gazzaniga，2000；Kolb & Whishaw，1990；Pinker，1997；Scarborough & Sternberg，1998；Sternberg，1999；Wilson & Keil，1999）。例如，即使像在咨询、学校、临床、心理治疗等应用领域，都从理论导向的基础研究中受益不少（Davidow & Levinson，1993；Fagley，1988；Garb & Schramke，1996；Lilienfeld，Wood，& Garb，2000；Mahrer，2000；Siegert & Ward，2002；Strupp，1989）。

可以想像一下古典和操作条件反射的基本行为原理，这些原理几乎完全是在非常人工化的实验室情境下，从鸽子、老鼠之类的、“非人类”动物身上发现的，然而这些原理已经成功地应用于许许多多的人类问题上，包括对儿童自闭症的治疗，大量数字材料的教学，酗酒、肥胖的治疗，精神病院

病人的管理，恐怖症的治疗等。这还仅是一小部分而已。

这些原理之所以可以做这些应用，是因为在实验室的研究中，研究者可以精确地去辨认外界刺激与行为之间的关系，这是在自然情境下无法做到的，因为在这种自然情境里许多行为之间的关系可能会同时起作用。至于用“非人类”的被试，许多例子已经证明，从实验动物行为研究中得出的理论是对人类行为规律的一个很好的初步仿真，当把人作为研究对象时所发现的规律及定理与用其他动物做实验所发现的结果非常相似。其实，这一发现在如今应该早已是不足为怪的了，因为人类疾病治疗的每一步进展现在几乎都是建筑在动物研究所收集的数据之上的（Kalat，1995）。例如行为治疗、减缓压力、心理治疗、受伤或残疾人士的康复、衰老对记忆的影响、对有神经肌肉障碍患者的辅助、药物对胎儿发育的影响、药物依赖、失忆、慢性疼痛的治疗等都是动物实验推进的应用实例（Domjan & Purdy，1995；Feeney，1987；Kalat，1995；Miller，1985b）。最近，一个在狗身上进行的研究马上要对认识人类焦虑症背后的原理做出突破性的进展了（Groopman，1999）。

实际上，“这不是真实生活”的批评已经导致了人们对动物实验的刁难，而这种刁难往往出于政治的原因。例如，那些为重度污染企业效劳的政客们就总是无视于在动物身上进行的有关污染是致癌的一个潜在因素的研究评估报告。他们宣称动物研究不可信，不能推广到人类身上，妄图为制造污染的企业开脱罪名。然而实际上，1988 年路易斯安那州的一组科学家在对 23 种致癌物（包括苯、石棉等）的研究中发现，

由动物实验所估计的死亡率与人类流行病学研究所取得的数据十分相近（Finkel，1996）。

心理学家有关知觉过程的研究在理论上也取得了瞩目的进展。他们从实验中推导出的规律和理论已经被应用于解决各种不同的实际问题，例如雷达监视、街道照明和飞机驾驶员座舱的设计（Nickerson，1992；Swets，Dawes，& Monahan，2000；Wickens，1992）。同样的，认知心理学家在研究人们处理和记忆信息的方式上也积累了大量可被反复验证的研究结果。虽然作为心理学中最年轻的理论，认知心理学在应用的范围上还不能和其他老牌理论相比，但是它的应用可能性已经在工业、教育等情境中得到检验。例如，在衰老对认知效果的影响方面我们已经积累了许多新的认识（Rabbitt，1993；Salthouse，1996；Salthouse & Miles，2002），而这些新知识有可能可以直接帮助我们设计对认知丧失者进行恢复训练的方案（Dixon & Backman，1995； Poon，Rubin & Wilson，1989）。

心理学家对判断与决策的研究已经应用于医疗、教育和经济等领域的决策制定（Belsky & Gilovich，1999；Davidow & Levinson，1993；Dowie & Elstein，1988；Fridson，1993；Gilovich，Griffin，& Kahneman，2002；Kahneman & Tversky，2000；Swets，Dawes，& Monahan，2000；Zweig，2001）。一个令人兴奋的新进展是，认知心理学在司法系统的作用逐渐受到重视，例如，信息收集时记忆的问题、证词的评估以及如何做判决等方面，都为检验认知心理学理论提供了很好的机会（Bornstein & Rajki，1994；Kassin，Tubb，Hosch，& Memon，2001；Koehler，1993；Kuhn，Weinstock，& Flaton，

1994；Loftus，1993；Wrightsman，1999）。另外，从20世纪80年代开始出现的许多阅读教学的理论与实践，也逐渐受到认知心理学的广泛影响（Ehri，Nunes，Stahl，& Willows，2001；Pressley，2002；Rayner，Foorman，Perfetti，Pesetsky，& Seidenberg，2001，2002；Stanovich，2000）。最后，在公众关注的热点问题上心理学家也给出了相当重要的科学依据，例如，要如何对待儿童在法律程序中提供的证词（Ceci & Bruck，1993a，1993b，1995；Ceci & Hembrooke，1998）以及受虐儿童所“恢复”的记忆是否是准确的（Bremner，Shobe，& Kihlstrom，2000；Clancy，Schacter，McNally，& Pitman，2000；Loftus，1994，1997；Pezdek & Banks，1996；Pezdek & Hodge，1999；Spanos，1996）。

正如前面提到的，认知心理学中的研究成果已经通过了“可重验性”这一基本检验。许多有关信息加工过程的基本原理已经得到全世界许多实验的验证和支持。但是很多人还是不了解，如果密歇根大学的一位心理学家得到了某个重要的发现，那么同类实验很快就会在斯坦福、明尼苏达、俄亥俄、剑桥、耶鲁、多伦多等多所大学以及其他地方进行重验。通过这种重验，我们很快就能知道这一发现到底是密歇根大学被试的特殊性所致，还是研究的实验情境所致。

当然，并非所有的心理学研究结果都能成功的重验（见Carroll & Nelson，1993）。恰恰相反，重验失败是有的，而且，我们往往从失败中得到教训和启迪要比成功的重验多得多。然而在认知心理学研究中，因被试的特殊性而造成的重验失败倒是很罕见的。实际上，重验失败的原因，大多是由

于实验刺激与方法上的细微差别所造成的。因此，失败后，通过仔细检查要在什么实验情境下，才能促发某行为现象，科学家可以从中更准确的了解这一现象，从而建立一个更精密的、有关其如何产生的理论。

或者有人可能要问，既然一些研究结果有重验失败的事例，那么它们还能被应用吗？又或者有人要问，当科学家之间还不能对理论的所有细节达成共识，那么我们把还没有得到定论的知识及理论拿去应用是否是太快了一点？像这一类的顾虑，在心理学知识的应用中经常见到。这是由于人们没有意识到，其他科学其实也是在它们的发现和理论还没有完全确立之前就开始应用了。当然，我们在第二章就清楚地讲过，任何一门科学都是要经过不断修正的。如果我们在应用科学的研究成果之前一定要绝对地确定它们的正确性，那么什么应用都不会发生了。所以应用科学家会尽其最大的努力去应用最准确的现有知识，但是，同时也意识到该知识还是有可能是不正确的。

“大学二年级学生”的问题

有许多人很关注心理学研究成果的“代表性”问题。他们的质疑主要集中在研究中所用的被试，而非实验设计的细节。然而，在心理学的许多领域，例如对基本感觉及认知过程的研究中，存在的争议多数是由实验设计所引起的。在基本的信息加工、脑的功能结构以及视觉系统的特征等方面，无论被试是蒙大拿州人还是佛罗里达州人（甚或是阿根廷人）其研究结果都是相似的。并且，这些结果与被试的父母是修补匠、裁缝还是教授都没有任何关系。

所有的科学都会假设一些因素对实验结果是不会造成什么影响的。生物学家一般不会担心培养细菌的器皿在厚度上之细微差异会对里面培养的细菌产生什么明显的影响。当然，这些差异有可能造成一定影响，科学的假设都不是绝对的，但是生物学家必须把他们的精力放在他们认为更有可能是影响因素的变量上面。这正是为什么，赫奇假设暗适应与人们的宗教信仰无关，所以他不去问他的被试是信仰基督教还是天主教。

在这里，我们还面临一个大学二年级学生（下称“大二生”）的问题，它是指参与心理学研究的被试中，有极大的比例，是大学二年级学生。由此问题引发了对心理学研究结果是否能推广到其他人群的质疑。心理学家们都很关注这一问题，因为它在某些研究领域中的确是一个严重的问题。然而，我们一定要只从一个角度来看这个问题，不要泛化问题。也就是说，并不是所有的研究都会因被试是大二生而出问题。更何况，心理学家对这一批评有几个合理的辩解。这里提出三点：

1. 大二生的批评并不会令研究结果不正确，只是促使我们开展更多的研究，以便可以对理论的可推广性进行评估。由于我们先前已经收集了大二生的数据，即使从其他人群中获得了相反的数据，而必须对理论作出相应的调整，也只会使理论变得更加精确，而不会完全否定它（Gage，1996）。即使在比较极端的情况下，结果完全被重验失败，我们充其量也只能说，建立在大二生所给出的数据之上的理论不够全面，而不能说其一定是错误的。

2. 在心理学的许多领域中，大二生并不构成问题。因为在基础理论（例如，视觉系统）的研究中，那些过程最为基本，没有人会认为他们的功能组织会受到样本的人口统计学特征的影响。
3. 研究在不同的地区取得一致的结果，可以较大地增加理论的地域普遍性，并在一定程度上，增加了社会经济、家庭、早期教育经验等因素的普遍性。例如，与50年前只有极少数精英才能读大学的大二生相比，现在的大二生已经是从各式各样社会经济背景来的人了，这与现在有许多的大学都已经平民化了有关。

然而，在心理学某些领域中，不承认大二生这一问题是不明智的。好在，心理学家正在努力纠正这一问题。例如，发展心理学家，由于其研究的内容，就必须非常关注这一问题。每年都会有数百名这方面的研究者，将众多用大学生做被试所得出的理论及发现，在其他不同年龄的被试身上进行重验。

从不同年龄群体所得到的结果与大学生的结果不会总是相同。对发展心理学家而言，如果老是得到完全相同的结果，那就太没意思了。因为他们就是要将年龄这一因素加入到心理学各个理论当中，以显示年龄的重要性。这一领域的研究也确保了心理学的宏大理论不是只是建立在从大学生那里收集的、微薄的数据基础之上。

发展心理学家也进行了一些跨文化研究，来验证只根据北美儿童所得到并建立起来的成长发展规律，是否具有全球普遍性。例如，史蒂文森等人（Stevenson et.al，1985）通

过给中国、日本和美国孩子完成的大量认知任务实验得出："三种文化下的孩子在执行这些任务时，所反映出的认知能力结构是十分相似的"（p.727）。其他跨种族、跨文化的比较也验证了史蒂文森等人的发现（McBride-Chang & Kail，2002；Rowe，Vazsonyi，& Flannery，1994）。在其他领域，也有许多跨文化比较的研究显示相似的趋势（例如，Day & Rounds，1998；Krull，et al.，1999；Rozin，Lowery，Imada，& Haidt，1999）。当然，也有不少跨文化研究并不能成功地重验用美国大学生所得到的研究结果（如，Menon，Morris，Chiu，& Hong，1999；Nisbett，Peng，Choi，& Norenzayan，2001；Peng & Nisbett，1999）。但当不一致出现时，这些研究结果仍然对某些理论及结果之受文化环境限制这一特点提供了重要的信息（Chang，1996；Choi，Nisbett，& Norenzayan，1999；Yates，Lee，& Shinotsuka，1996）。

教育心理学家也考察了大二生问题。例如，他们与发展心理学家及其他教育问题研究者合作，共同制定了一种测量基本认知能力的测验，用于预测学生未来的学业成绩，例如阅读掌握率。这种测验就有一定准确度，并且这一准确度不受学校所在区域、社会经济水平以及种族等因素的影响。

对大二生以及代表性问题的批评，多是针对社会心理学而来的，因为这一领域经常使用大学生样本，通过实验室研究，以发展有关社会交往、群体行为以及社会情境中的信息加工过程等方面的理论（Kunda，1999）。然而，即使是在这一心理学领域，也有许多的证据可以显示，在实验室里发

展出来的关系及理论还是能够预测在许多不同情境中不同类型人的行为。

例如，几年前，威斯康星大学的心理学家里昂纳德·伯克维兹（Leonard Berkowitz）曾提出了所谓的“武器效应”。它是指个体的环境中如果有武器出现，就会增加个体的侵犯行为。这一发现最初是在实验室中进行的，可以说是不具代表性的典型实验。所以这一结果受到了猛烈的抨击，认为研究结果是在特别设计的情境被诱导出来的行为。然而，之后针对这一理论所做的研究都不能推翻它，无论研究者用的是什么工具来测量侵犯行为；无论是在欧洲还是在美国；无论被试是儿童还是成人；无论是在实验室之内，还是在现实生活情境中当参与研究的被试根本没有意识到自己是在做实验（Berkowitz & Donnerstein，1982；Turner，Simons，Berkowitz，& Frodi，1977）。有些研究者甚至将武器效应背后的认知机制给提取出来了，认为是一个语义记忆被（武器的出现）自动启动的过程。这个效应是如此强烈，令安德森等人（见 Anderson，Benjamin，& Bartholow，1998）把他们的研究报告的题目定为：“是手枪自己扳动了扳机吗？”

认知、社会以及临床心理学家都对人类决策行为展开了许多研究。最初，大多数的研究都是在实验室中进行的，并多用大学生作为被试，让他们完成一些极其“人为化”的任务。然而，从这些研究中所得到的关于决策行为的许多原理，已经在大量的非实验情境中被成功地重验，这些情境包括银行家对股票价格的预测、纸牌赌博游戏、精神病医师对病人行为的预测、经济预测、军事情报分析、全美橄榄球联赛的博彩、技

师对维修时间的估计、房地产经纪人对房价的估计、商务决策以及医师的诊断等等（Bazerman，1999，2001；Braun & Yaniv，1992；Davis & Holt，1993；Goldstein & Hogarth，1997；Hammond，1996；Heath et al.，1994；Tassoni，1996；Wagenaar，1988）。并且这些原理还在相当实用的个人理财咨询领域中也得到了应用（Bazerman，1999，2001；Belsky & Gilovich，1999；Kahneman & Tversky，2000；Thaler，1992；Zweig，1998，2001，2002）。

博恩鲍姆（Birnbaum，1999）预计互联网将为心理学的大二生问题提供一个解决办法。他曾在实验室中作了有关决策问题的一系列实验，并在互联网上征集了一批志愿者来参与。结果发现在实验室中用大二生的研究结果，与后来在网上征集来的被试所给出的结果相同，而后一批被试来源要比前者广泛得多——包括了来自44个国家的1224名志愿参与者（见McGraw，Tew，& Williams，2000）。

仅从这少数几个例子，我们已经可以看到心理学研究结果的稳定性及普遍性经常被人低估了（见Cheng，2001；Gage，1996；Lipsey & Wilson，1993；Rosenthal，1990）。安德森，林赛和布什曼（1999）曾报告了他们对实验室研究与田野实地研究结果的比较研究。在对包括许多不同课题，如侵犯行为、领导行为和抑郁症等的大量研究进行考察之后，他们得出来一个相当一致的结论：不同的研究情境并没有带来不同的结论，反而总是带来相同的结论。

“真实生活”与“大二生”：从一定的角度来看问题

本章中提到了几个焦点问题，在这里和大家一起回顾一下，以便弄清楚哪些是我们说过的，哪些是还没说的。我们谈到了由于人们对科学研究的原则，存在着一些误解，以致使他们对心理学研究的“人为性”经常提出一些批评。我们也谈到了为什么人们会关注心理学家没有在他们的所有研究中使用随机取样的样本这一问题，以及为什么说这种担忧是没有必要的。最后，我们谈到了一个较为合理的批评，即大二生的问题，但这个问题也常常被过分的夸大了，尤其是那些对心理学研究的多样性和广泛性不熟悉的人（见第一章）。

尽管如此，心理学家确实应该对他们的实验结论谨慎处理，不要过分依赖一种方法，尤其是同一类被试群体。在下一章中，将对这一点进行讨论。事实上，心理学中确实有一些领域深受“大二生问题”之苦（Wintre，North，& Sugar，2001）。而作为这一问题的解药的跨文化心理学，则尚未发展成熟。然而，我们仍然对这个问题的缓解抱有很大的希望，因为在心理学研究者的行列中，自我批判受到尤其的重视（见第十二章；Anderson & Gunderson，1991；Dawes，1994；Heinsman & Shadish，1996；Kimble，1999；Proctor & Capaldi，2001；Redding，1998；Wintre，North，& Sugar，2001）。实际上，已经有许多著名的心理学家以“批判”为其本职工作（Leary，2001；Robinson，2001）。并且每年各类科学杂志上，都会出现就心理学研究所用方法上的漏洞，或是对大二生问题提出警告。而心理学界已经对后一

问题非常关注。所以，我们当然不应该忽视它，但是我们也必须站在一定的角度来看待它，不要将之泛化。

小　结

有些心理学研究是应用性的，其目的是将其研究结果直接应用于某些特定的情境中。在这类应用性研究中，结果是被直接推广到现实中的自然情境里，因此样本的随机取样性与情境的代表性相当重要。然而，心理学的大多数研究却不属于这一类。它们多是基础研究，是用来验证人类行为背后机制的理论的。在这类的基础研究中，被试是不是随机取样及研究情境是不是有代表性，并不是特别重要，因为这些研究关注的重点是验证一个理论其预测的普遍性。事实上，在验证理论的基础研究中，正如前一章说明的，人为的情境是故意设置的，为的是要将研究中的关键变量分离出来，并且用之来控制那些无关紧要的变量。从这一个角度视之，心理学实验的“不是真实生活”，恰好是它的长处，而非其短处。

（郑智敏　译）

第八章　别了，爱因斯坦综合征

——证据集中的重要性

“生物学实验披露生命的奥秘”，“思维控制获重大突破”，“加州科学家发现延缓死亡的秘诀”——对以上这些形形色色、充斥于小报头条的所谓“突破”性的新闻，可谓屡见不鲜。正因为这些新闻通常来源于那些最不负责的媒体，所以大多数科学家都建议公众要用一种怀疑的态度来对待这些新闻。这一警告当然是应该的，也不足为奇。不过，本章的目的，不仅仅是告诫公众有关这些夸大而散播出去的错误信息，或是警告读者，在评价有关科学进展的报告时，你必须审慎地查核消息的来源。这一章中，我们还想引入一个比前几章中所讲述的、更为复杂的科学进展观。为此，将详细阐述第一章中已经介绍过的系统实征主义和公开的知识这两个构想。

媒体常用的“突破”作为大标题，其实是误导了公众对心理学和其他科学的认识。一个特别典型的曲解就是，它们诱使公众认为通过一个关键性的实验，就能解决某一科学研究领域的所有问题；或者说，科学的进展可以由于一个单一的、具有批判性的新见解一蹴而成，而这一新见解能够完全推翻先前众多研究者所累积的所有知识。这一广为流行的科学进展观倒是与新闻媒体的操作方式十分吻合。在媒体那里，历史往往是用一些支离破碎的、互不联系的事件连串起来的。

此外，这一科学观也与好莱坞娱乐业的运作方式相近，好莱坞的电影剧本必须有一个事件开端和一个令人满意的结尾，令人满意是指不能留有任何的悬疑及不确定因素。然而，这一观点实际上只是对科学进展的一个相当粗糙的描绘。说得更严重些，它导致了人们对科学进程的曲解，并减弱他们在某一个问题上评估科学知识的能力。在本章中，讨论科学研究的两条原则——关联原则（Connectivity Principle）和证据集中原则。与“突破式”模型的进展相比，这两条原则能够更准确的描绘科学进展。

关联原则

尽管我们不认同科学进展取决于“大跃进式”或“关键性”实验的观点，但并不是说，在过去从来没有过关键性的实验和理论上的大跃进。相反地，在科学史上一些最著名的实例都证明了这一类的实验及跃进的确存在。其中最有名的，可能就要数爱因斯坦提出的相对论。这里，我们看到让人惊叹的理论洞察力导致人们最终重新构想诸如时间、空间和物质等基本概念。

然而，正是爱因斯坦所取得的这一辉煌成就，使之成为公众心目中关于科学进展的主导模型。这一模型之所以会一直延续下来，主要是因为它与媒体常用来报道新闻的隐含“脚本”高度吻合。人类历史上还没有哪一个观念遭遇像相对论那样的麻烦：那么多的胡言乱语，那么多的不实推论都紧随其后（要知道，爱因斯坦并没有证明“一切东西都是相对

的”——参见 Holton，1996）。当然，我们的目的并非是去纠正这些谬误，但是这其中倒可以帮助我们在讨论如何评价心理学理论时带来一些启迪。

爱因斯坦理论引发人们对物理世界最基本观念的重新认识，是这样的基础及重要，以至于在许多通俗读物中，常常将它看作是类似艺术中的观念转变（就像一位二流诗人被重新评价成一个诗坛天才；一种艺术流派被宣告死亡等等）。然而，这一比较恰恰忽略了艺术与科学在观念转变中所存在的基本差异。科学研究中的观念转换遵从关联原则，而这一原则在艺术中并不存在，或至少说是极为少见的（见 Bronowski，1956，1977；Popper，1972）。

一个新的科学理论必须与先前已建立的实征数据相挂勾。只有在解释新现象的同时又能很好地兼容原有的事实证据，这一理论才会真的被认为是一个进展。新的理论可以用与先前理论截然不同的观点来解释旧证据，但是它必须能解释得通。这一要求确保了科学能够不断地积累、向前发展。除非一个理论的解释能力所适用的领域能够被大大拓宽，否则不会有真正的进步。如果一个新理论能够解释新观察到的现象，但却无法与大部分的旧事实兼容，那么人们不会承认它是对旧理论的全面更新，也因此不会立刻用它替代旧理论。相反地，在这种情况下，旧理论和新理论会在观念的角斗场上继续搏杀，直到有一个可以整合两者的新理论出现，从此这两个较力者都被淘汰出局。

尽管爱因斯坦的理论带来令人叹为观止的全新观念（运动中的时钟走得比较慢，质量随速度增加），它们仍然遵从关

联原则。它虽然抛弃了牛顿式机械力学，但是它没有否认或抛弃牛顿理论的基础，即有关物质运动的实征证据。恰恰相反，低速运动时，这两个理论做出了基本相同的预测。爱因斯坦体系的高明之处在于：它能够解释与牛顿力学不兼容的众多新发现的，有时甚至是令人完全意想不到的现象。这样，我们看到，即使是爱因斯坦的理论，这个科学史上最惊人的、最基本的概念重构，实际上都遵从关联原则。

消费者鉴别好坏的方法：看看是否违背了关联原则

科学进展的“大跃进式”模型——我们姑且称之为“爱因斯坦综合征”，经常将我们引入歧途，因为它让我们觉得新发现必定要违背关联原则。这一观念是非常危险的，因为一旦放弃了关联原则，主要的受益者将是那些伪科学和虚假理论的传播者。这些理论的动人之处和公众魅力部分来自于：它们常常披着令人惊叹的新外衣。“你看，相对论不是也在当时被认为是全新的吗？”这一辩词已经变成一种策略，试图将新奇转化为优势。当然，一旦伪科学家们试图打入某些科学领域，这些领域先前所积累的大量数据就成了主要障碍。实际上，这一障碍也是微不足道的，因为伪科学家们还有两个强大的法宝来应付这一障碍。在第二章中曾经讨论过其中的一个法宝，即在解释先前数据时，使自己的理论永远无法验证它是错的，因而就令先前数据变得毫无用处了。

否认先前实征证据的第二个法宝是宣称这些证据不相关。这种否定通常都会着重强调其新理论的突破性。因此，常常看到类似“对现实的新观点”和“全新的突破”之类的用

语。但实际上，其花招还在后面。因为接着，这一理论会被认为是这么的崭新，以致使依据其他相关理论所获得的实验证据，都被认为是无关的，而只去考虑根据这一新理论的构念框架所收集的支持性数据。这就使关联原则被彻底地打破了。很明显的，由于理论是全新的，以至于他们可以理直气壮地说：与之相关联的实征证据尚不存在。这样，一切都没有问题了：伪科学拥有了一片能够使其蓬勃成长的沃土。那些旧的、“不相关的”数据都没有了，那些新的、相关的数据还没有做出来。正是由于“爱因斯坦综合征”让人忘记了关联原则，这才使得这种诡计得逞。然而，颇具讽刺意味的，恰恰是爱因斯坦的理论本身论证了关联原则的重要性。

加州大学的古生物学家凯文·帕典（Kevin Padian）举出另一个例子来说明，当关联原则的重要性被忽视时，公众将如何曲解科学的本质。堪萨斯城教育委员会曾决定把进化论排除在学生的必修课程之外，帕典针对此事，指出：“可见对科学如何进行整合人们完全缺乏理解……那种认为可以把一部分科学——特别是像进化论这样一个可以把整个生物学组织整合起来的主要理论——捆起来丢掉，而不会产生任何坏影响的想法，实在是太荒唐了”（Carpenter，1999，p.117）。生物哲学家鲁斯（Ruse）注意到进化论与许多独立、相关不大的科学领域都具有关联性，这些领域包括古生物学、胚胎学、生物形态学、生物地理学、神经科学等等（Ruse，1999）。与此类似，赫门（Shermer）指出“假设宇宙和地球只有大约1万年，那么当代科学的知识，例如宇宙论、天文学、物理学、化学、地理学、古生物学、古人类学及

早期人类历史学等都将是错的”（Shermer，1997，p.143）。知名的科普作家和古生物学家古尔德（Stephen J.Gould）表示，“失去进化论的生物学教学，就如同教授英语但不教语法一样”（Wright，1999，p.56）。

鲁斯（1999）曾谈到达尔文如何善用关联原则的一个例子。当达尔文的新理论与其他学科之间没有必需有的兼容性时，他就断然放弃了那个新理论。当时，达尔文想寻找一种遗传机制，与他的自然淘汰理论相匹配。他试图建立一套泛生理论，认为“身体的各部位都会给出一些很小的胚芽，这些胚芽在体内循环，并在性器官聚集，从而将之传送到下一代”（p.64）。该理论的一个问题就是它与细胞理论不相符合。另外，达尔文也无法解释这些胚芽是如何传送的，因为血液输送实验已经证明它们是无法通过血液传送的。由于这个以及其他种种原因，泛生论最终“因为与生物学里的其他领域不相兼容”（p.64），而在科学阵营中云消烟灭。

心理学也一样。哲学家马丽欧·邦格（Mario Bunge）曾提出，如果认知心理学一开始就否认经典和操作性条件反射的存在的话，它绝对不可能在心理学领域中立足，因为它就与行为科学中的其他知识无法建立关联了（Bunge，1983）。回想在第六章对“受辅助的沟通技术”的讨论，之所以怀疑它治疗语言障碍的功能，是因为它违背了关联原则。如果它是正确的，即意味着要推翻所有在神经、遗传及认知心理学等相当多不同性质的科学领域中所累积的基本原理。这一治疗的理论假设与科学中的其他知识没有任何关联。再想一下反对进化论的神创论也是如此，它的问题也在于它与科学中的其他知

识毫无关联，无论是生物学、地质学、生态学、化学或是遗传学。而进化论则不同，正如生物学家杜让斯基在一篇著名的文章中所说的：没有进化论，生物学将没有任何价值。进化论与其他科学有着密不可分的关联（Dobzhansky，1973）。

"大跃进式"与"逐渐整合式"的进展模型

将爱因斯坦式的革命当作是科学研究的典型，会诱导我们认为，科学进步依赖于许多的大跃进。这种"爱因斯坦式综合征"并非完全凭空捏造。在许多科学进展中，的确存在着在概念理解上的迅速及突然的大跃进。但问题在于，人们总是倾向于把这些例子泛化成科学进步理所应当的形式。实际上，许多科学领域所取得的进展，并非依靠单一的突破，而是靠一系列难以名状的、停顿及前进的交互更迭。

一般人通常并不察觉在科学的工作中所涉及的模糊性，科学实验很少能够完全解决某个研究问题，亦即在支持一个理论的同时可以排除所有其他可能解释同一现象的理论。新理论极少会很明显地优于所有先前存在并与之相竞争的构念体系，某一特定问题的解决通常不像科普电影里描绘的那样，是靠某个关键实验一蹴而成的，而是要等到科学界逐渐开始有了共识，认为支持一种理论的证据比支持其他任何理论的要强很多的时候。科学家们用来评估的证据并非源于单一、设计完美的某一个实验。相反地，科学家们常常必须对几十篇实验论文的结果进行总体评估，这些实验中每一个都有缺点，但又都提供了解决问题的部分答案。科学进程的这种更迭模式常常不为

人所知，是因为爱因斯坦综合征在人们的心目中造成了深刻印象，认为所有科学都与物理学一样，而不清楚物理学是出现大跃进式科学进展模型最多的学科。

试想一下，近几百年来，遗传及分子生物学的突飞猛进，这些进展的获得不是靠一个爱因斯坦式的天才，在一个时刻就把事情搞清楚的。生物学的现代整合是建立在无数次、各有缺陷的实验研究之上的；这些进展并不是靠取得对一个重大创新构念的认可就可以了，而是在几种得到支持的可能解释之间长期、反复的拉锯中逐渐形成的。经过十几年没有定论的实验、无数次的理论推论、辩论与批评，科学家们终于弄明白基因到底是由蛋白质还是由核酸组成的。观点虽然改变了，但是它绝对不是在一次大跃进中完成的。原子核的发现者欧内斯特·卢瑟福（Ernest Rutherford）在强调关联原则的重要性时，也指出“科学家们依靠的不是某一个人的思想，而是成千上万个人的智慧结晶”（Holton & Roller, 1958, p.166）。

卢瑟福的观点描述出另一个区分科学和伪科学知识的方法。科学是一种累积，它遵从的是关联原则，其特性之一就是有许多人的参与，而我们是以这些人能帮助我们理解自然界的程度来决定他们的贡献。没有任何一个单一的人，可以凭借其特殊的地位而控制科学的进程。科学不允许只有少数人才能享用的特殊知识的存在。实际上第一章中已经讨论了科学的这一公开性。相反地，伪科学则常宣称某些权威和研究者有某种“特殊”的渠道，可以比别人更容易找到真理。

在这里讲述了两个观点，它们提供了帮助我们理解心理学

这门学科所需的背景。首先，科学中不存在任何设计得十全十美的实验，任何一个实验，对它数据的解释都存在一定的模糊性。科学家经常不会等到完美或关键的实验出现后才去评估一个理论，而是将评估建立在大量有些缺陷的实验所呈现的整体趋势之上。其次，许多学科尽管没有爱因斯坦那样的人物出现，仍然在向前发展。它们的进展是以动静更迭的方式，而不是通过爱因斯坦式的伟大整合、分段跃进的方式来进行的。同样，心理学和许多其他科学一样，也是由众多不断增长的知识碎片聚集而成，没有一个整合的主题。

证据集中：在缺陷中进步

先前的讨论引出了在心理学中，在作证据评价时至关重要的另一个原则。这个原则常被称为是证据集中原则（或操作集中原则），科学家和科学应用者经常需要去判断，大量的研究证据究竟在说什么。在这种情况下，证据集中原则成为一个重要的工具。对使用科学信息的外行人来说，证据集中原则也是很有价值的，而且在评估心理学的研究成果时特别有用。尽管对证据集中原则展开全面技术性讨论将会偏离本书的主题，但有关这个概念的实践应用却是很容易理解的。下面，探讨一下这个原则的两种表达途径，一种称为缺陷实验的逻辑，另一种称为理论检验。

从极端观点看，一个实验可能存在无数个可能出错的可能性（或用专业术语来说，就是变得混淆不清）。然而，在大多数情况下，不会存在数量很多的关键性混淆因素。对一

个特定问题，较有经验的科学家通常很清楚，大多数的关键因素是什么。所以当考察实验证据时，科学家们通常知道每一个具体实验的关键性缺陷是什么。于是，证据集中原则提示我们去审察相关的研究文献所呈现的缺陷类型，因为这类研究的性质要么支持，要么削弱我们想要得到的结论。

例如，假设从一系列不同实验得到的结果都一致性地支持某个特定结论。既然这些实验都有缺陷，我们可以对这些研究存在缺陷的程度和特点进行评估。如果所有这些实验都拥有相同的缺陷，这一情况将削弱我们对所得的研究结论的信心。因为人们会假设：研究结果所获得的一致性，是由所有实验共有的某个缺陷所造成的。另一方面，如果所有实验的缺陷各自不同，那么我们对实验结果就比较有信心，因为看上去，研究结果的一致性不致于是由某一“污染”因素而令所有实验结果变得混淆不清。正如两位安德森（1996）所指出的：“不同的方法似乎孕育不同的假设。当建立在不同假设基础上的某一概念假设能够通过众多的、基于不同假设的证伪检验时，我们可以说是得到了一个有说服力的结论”（p.742）。

每一个实验都能帮助我们校正其他实验设计中存在的错误，同时它自身又在接受其他研究对其缺陷的检验中得以巩固。大量的实验，尽管各自多少都存在不同的缺点，并且实验技术也各有优劣，但只要能得到相近的结果，那么我们可以说我们的实验证据得到集中了。尽管没有一个实验设计会十全十美，但我们还是得到了一个强有力的研究结论。于是，证据集中原则要求我们，必须将研究结论建立在一系列彼此略微不同的实验所得到的数据基础之上。这个原则使得我们可以

做出比较强有力结论，因为由这样的一个情况所得到的结果一致性比较不可能是由某一个特定的实验程序所造成的。

证据集中原则也可以用理论检验这一方式加以论述。只有当一系列实验一致地支持了某一理论，而又集体地排斥了其他重要的竞争理论时，我们才说这一研究领域是高度集中的。尽管没有单个实验可以排除所有可以替代的其他解释，然而如果可以收集一系列实验，把它们放在一起看，都呈现了某一种趋势，也可以让我们得到一个比较强有力的结论。

最后，引入证据集中原则有助于摒弃另外一种错误观念。这种观念可能是由于在第二章我们讨论可证伪性时的过分简化造成的。那里的讨论似乎让人感觉，当第一个与自己的理论不相符的证据出现时，这个理论就算是被证伪了。然而，事实并非如此（Pigliucci，2002）。正如理论是被集中了的证据所支持一样，它也是被集中了的研究结果所否定。

心理学中的证据集中

强调集中重要性的原因在于，心理学的结论通常要靠集中了的证据来确定。这一事实当然并非独特或与众不同（许多其他学科的结论也都不是建立在单一的、具有决定性的实验基础之上，而是要靠对几十个模糊实验的整合）。但是我们有理由认为，这种情况在心理学中尤为突出，因为心理学实验通常鉴别事物原因的能力都比较低。也这就是说，支持某一理论的数据往往只排除一小撮可替代的可能解释，同时保留许多其他可能的理论作为候选。其结果是，只有收集大量研究数据后，才能得到较为集中，因而较为有力结论。

心理学实验具有高度不确定因素，这一点毫不令人惊讶。因为其研究问题涉及人类行为，具有巨大的复杂性，如果心理学家们能公开承认这个点，并且致力于向公众解释这个事实所带来的后果，公众可能会对心理学这门科学有更多的理解。心理学家们应该承认，心理学是一门科学，并且正处在发展之中，虽然它的发展很缓慢，并且许多结论往往都来自日以继夜、相当痛苦地统合与争论中。对媒体经常宣称的所谓突破性进展，我们要永远带着怀疑的态度，这点在心理学领域尤为重要。

在心理学这一领域，我们必须格外小心。例如，我们必须抗拒将一个周围实验证据尚很模糊的心理学假设作为是已经“证实”的理论。本书的好几章里都在强调大家必须抱有这种怀疑态度。要注意不要从相关中推论因果，以及拒绝接受见证叙述式的证据等，都是例子。与此同时，也不要因为知识不完整或结论还有待进一步探索就妄自菲薄，开始怀疑心理学究竟能否产生强有力的结论。更不可因此就做出不理智的结论，认为心理学永远也无法成为一门科学。从这个立场出发，证据集中原则可以看作是，前面所提到的，不要过度去解释尚不成熟的研究结果这一警告的一个平衡砣。尽管心理学研究有许多缺陷，但是不要气馁，各式各样研究逐渐集中到一个方向，还是能够得到合理、有力的结论。

观察证据集中原则最好的方法就是去检验存在问题的心理学领域，这些领域往往通过应用了集中原则而得到结论。让我们再看一个例子。关于电视暴力节目是否增加儿童攻击行为这一课题能充分说明证据集中原则的重要性。目前，对这个

问题，科学界的认识较为一致：儿童观看电视暴力节目，确实增加了攻击行为发生的概率。这个效应虽不是非常大，但确实存在。科学家们对这结论的信心，是通过分析上百个不同的研究结果，而不是在一个具有决定性的研究中得到的（如，Anderson & Dill，2000；Anderson，Huston，Schmitt，Linebarger，Linebarger，& Wright，2001；Bushman & Anderson，2002；Friedrich-Cofer & Huston，1986；Huesmann & Eron，1986；Paik & Comstock，1994；Pearl，Bouthilet，& Lazar，1982；Wood，Wong，& Chachere，1991）。这些研究所用的实验设计、被试群体及具体的技术都非常不同。我想现在大家一定已经清楚了，这些差异恰恰是这个领域的力量，而不是缺点之所在。

电视公司的老板们自然不愿意看到自己充满暴力的节目对儿童存在负面影响的证据。于是，他们就发起了打击这些研究是在"误导公众"的宣传攻势，他们利用的正是公众没有认识到的这一点：即研究结论是建立在许多研究的集中证据上，而不是哪一个关键性、决定性的研究上（因为公众正是"爱因斯坦综合征"的受害者）。电视公司不断挑出某一个研究来加以批判，并且暗示通过论证每一个研究都存在缺陷，就可以否定相关的总体结论。尽管社会科学研究者可以一一去响应对这些研究的批判，但是大多数时候，研究者是很乐意承认这些研究是有缺陷的。这一点，是公众没有认识到的。这是因为，本着证据集中原则，研究者并不认为承认某个特定研究的缺陷，就否定了关于电视暴力影响攻击行为所得到的一致结论。这表示，在支持一个结论的众多研究中，也许某

些有着这样那样的缺陷，但是另一些不带这些缺陷的研究，也得到了同样的结果。这些研究自身可能有问题，但是另一些研究则等于纠正了这些缺陷，也得到了类似的结果。

例如，有关电视节目与暴力行为关联问题的早期研究揭示，儿童观看暴力节目的数量与暴力行为之间存在相关。当时就有人很正确地指出，这种相关证据不能被看作是因果结论。也有可能是第三个变量造成这一联系，或者攻击行为高的儿童倾向于看比较多的暴力节目（方向性问题）。

然而，科学界并非仅依赖上述相关证据就会做出任何结论。我们也可以用许多比简单测量两变量之间关系更复杂的相关分析技术，这些技术允许我们做出一些探索性的因果推论（第五章中提到的偏相关就是其中之一）。这些技术其一就是，采用纵向的研究设计，在不同时间点测量相同的两变量——在这里就是电视暴力和攻击性。由这一设计所得到的相关模式可以告诉我们两者是否有因果联系。这一类型研究已经有人做过了，得到的结果表明：观看暴力电视节目的确有增加人们日后攻击行为的可能性。

也许，这时有人又会质疑起纵向相关技术，认为它是具有争议性的。这也确实是如此。但是，最重要的一点是，电视暴力和攻击行为之间有因果关系的结论并不完全依靠相关证据，不管是简单相关或是像纵向相关技术这样复杂的相关分析。实际上，已经有许多研究是用实验法来进行验证工作的。在这些实验中，大量电视暴力被研究者操纵，而不是仅仅对已有的节目（现实中自然的情境）进行暴力评估。第六章中讨论过，在其他变量受到实验控制的同时，只对一个变

量加以操纵，就可以避免相关研究对因果及第三变量问题的困扰。如果其他变量均完全相同的两组儿童表现出不同的攻击行为水平，而且如果这两组仅有的差异就在于：一组看了更多暴力节目，而另一组没有；那么我们就可以比较放心地推论说：被操纵的变量（电视暴力）引致结果变量（攻击行为）的改变。在大多数相关的实验研究中都得到了这一结论。

这些研究当然又引来了另外一些批评，提出了“这不是真实生活”的抗议。在前一章中已经讨论过的这一抗议的论点，以及利用这些论点所进行的恶意攻击。无论如何，有关电视暴力影响的研究结果并不是只在一群很古怪的小孩子身上才发现。事实上，这一结果在美国不同地方，以及世界的不同国家都成功地得到反复验证。使用不同的实验情境、不同的电视节目作为研究对象的各个实验，也都得到高度一致的结果。

更重要的是，不仅在实验室中，就是在现场研究，大都也得到了相同的结论。现场实验（*field experiment*）的研究设计，也被用来探讨电视暴力 - 攻击行为这一课题。这类研究设计的发明提醒我们，避免惯性地认为实验设计和实验室之间的必然联系。人们有时候认为我们只有在实验室里才能操纵变量，在现场研究中才能探索相关。这个想法是不正确的。实验室里也常常可以进行相关研究，而非实验室的情境下也常常可以操纵变量。在非实验室的情境下操纵变量进行现场实验研究，尽管有时需要相当大的创意，但却在心理学领域中变得越来越普遍了。

当然，现场实验研究自身也存在缺点，但是这些缺点往往正是其他研究设计的强项，所以可以相互弥补。总而言之，支

持电视暴力导致儿童攻击行为增加的这个观点的证据不仅依赖于某一个特定的研究，或某一类特定的研究方式。

这个情况类似于吸烟和肺癌的关系研究。烟草公司老板常常试图迷惑公众，暗示公众：香烟引起肺癌的结论只存在于某些特定的研究中，并且攻击这些研究站不住脚。然而，大量证据都集中指向一个强有力的结论。来自若干不同类型的研究，其数据都趋向集中于一个结论，而且不会因为单一实验受批评而发生重大改变。

在这里，很适合我们讨论一下类似肺癌病因之类的医学问题。在医学诊断和治疗中，许多决策都建立在不同类型研究结果是否集中得出一个结论的基础上。例如，当流行病学（可以说是一种涉及人类的现场研究，目的在于寻求某一疾病与现场中其他环境及地理因素的关联），精确地控制的动物实验，以及人类被试的临床试验等不同类型的研究结果，都趋向集中于一个结论时，医学界才会对这一结论具有比较大的信心，认定这一结论是可靠的，而医生们才愿意在这些证据的基础上开展治疗方案。

然而，上述这三类研究都各自有其缺陷。流行病学研究通常是相关研究，变量之间存有虚假联系的可能性很高。实验研究虽然具有高度控制，但被试往往是动物而非人类；临床试验虽然是在医院环境下，将人类被试放于一个真实的治疗环境之中，但是又因此而有许多因素没有受到控制，例如安慰剂效应以及治疗小组对病人的治愈期望等。虽然这些不同的研究设计都各自有其问题，但是，当用这些不同的方法所得到的数据都相当紧密地集聚在一块儿的话，医学研究者有理由

去肯定这个高度集中的结论。就像心理学家可以用这一证据集中原则来帮助他们对诸如电视暴力对攻击性行为之影响的研究，做出强有力的结论一样。

科学研究的共识

对电视暴力影响的评估是一个很典型的例子，告诉我们如何通过数据的积累，最终回答心理学所提出的问题。特别是当涉及到一些迫切需要解决的社会问题时，更需要牢记这些问题的答案不会是顺手拈来的，总要累积许多不同实验的结果，而绝不可能因为出现一个突破性的研究就一下子把问题完全解决。用一个简单道理来总结：在评估心理学的实征证据时，心中要想的是"科学共识"，而不是"重大突破"，是"逐渐集中"，而不是"飞跃进步"。

遵守"科学共识"，而不是"重大突破"原则，在围绕着"幼童补偿性教育计划"的种种争论中，起了很重要的作用。在20世纪60年代后期以及70年代初期，人们曾就林登·约翰逊总统要建立一个"优质社会"所提出的一系列社会政策项目是否真的有效，展开了许多争论，在这一争论进行得如火如荼之际，公众经常会在报纸上看到这样的头条:"早期介入可将智商提高30分"和"心智开启项目是一个失败"等等。一个外行人该怎样面对如此矛盾的信息呢？在这个例子中，"科学共识"而不是"重大突破"原则，无疑可以提供一些帮助，因为这一原则提醒我们，这两个报纸头条所指的实征结论可能都言之过早。实际上，也确实又花了十年的

时间，研究者才对这一重要的社会问题达成科学研究的共识。

这一共识不是建立在单一的、关键性的研究基础之上，而是建立于康乃尔大学一组专家的精心分析工作之中（Lazar, Darlington, Murray, Royce, & Sniper, 1982）。在进行这一分析时，这组研究者合并了20世纪60年代到70年代早期收集的包括11个幼儿教育项目所提供的成百上千个被试的数据。尽管单一研究项目的结果有时难以解释，但当把它们汇聚起来之后，整体结果却还算清晰。早期介入的短期教育项目并不一定会将智商提高30分。同样，“心智开启”和诸如此类的项目也并非绝对失败。早期教育介入项目确能对参与此项目的儿童的后期教育产生具体的、可靠的影响。这些儿童留级的比较少，被送入特殊教育班较少，对学校和学业成绩有比较多的正面态度，并且表现出持久的学业进步（参见Lee, Brooks-Gunn, Schnur, & Liaw, 1990; Bamey, 1999）。

加拿大心理学家提摩西·摩尔（Timothy Moore）曾讨论了如何通过增加人们对证据集中原则的敏感度，来促使法庭审判更好地应用专家证词（Moore, 1996）。他特别讨论到，当某一个专家的证词只是表达个人意见，难以代表该领域专家的共识。摩尔给出了一个叫犹大教士（Judas Priest）的案例。这个案例是关于两个少年自杀事件，他们的父母将“犹大教士”的乐队告上法庭，因为他们认为是该乐队所奏出的音乐隐含着某些潜意识的信息，诱导了他们的儿子去自杀的。尽管当时科学在这一领域的共识是，没有任何证据显示那些潜意识信息会产生导致人走向自杀的效果（即使现在，这一共识仍然成

立）。然而该案的法官显然受到某一个坚信用心理动力学来解释这一事件的学者的影响，而没有听从科学界大多人的共识。摩尔就此下结论说：

但是，我们最终的分析结果是，误导法庭的不是那些一眼就可揭穿的伪科学，而是这位专家的误导……他的观点虽然极富想像力和逻辑性，但是却与先前科学界对这个现象的理解大相径庭。很出色的履历和出众的职位并不足以保证一个专家的观点具有科学依据。任何一个专家，如果他的证词独树一帜，风格奇特，且不为科学界的多数同行所接受，那么他就没有在法庭上发挥他应发挥的作用——将他的专业知识传授给法庭。（p.38）

方法和集中原则

集中原则也告诉我们，心理学所有的研究领域都应该使用多种研究方法来研究同一个问题。由于不同类型的研究技术各有千秋，从各种方法所得到的结果之中求取一个平衡作为一个问题的结论，是比较妥当的。长久以来，心理学一直被批判为过分依赖实验室的实验技术。这种批评也是心理学家们相互争论的一个话题。它是不是正确要看具体是在讨论哪一个特定的研究领域。但是，不可否认，心理学的各个领域近年来都已经开始使用不同的研究方法了。例如社会心理学家，他们曾因为过于依赖实验法而受到最多批评，而如今都已经开始大量转向富有想像力的现场设计，从而可以依藉证据集中原则来支持其理论（Kunda，1999）。

心理学家比伯·拉坦（Bibb Latané）和约翰·达利（John

Darley）的工作提供了一个很好例子。他们因研究所谓"袖手旁观"现象而广为人知。这个现象是指，在紧急事件发生时，人们不向受害者施助的现象。拉坦和达利（1970）发现，在许多紧急事件发生时，旁观者会向受害者施助的可能性，会因为有其他旁观者在场而降低。

然而，这两位研究者清楚地知道，如果仅依靠观察实验室中的被试，在遇到紧急事件时的反应，来下他们的结论，那么这些结论就太令人置疑了。因为在实验室中，目睹紧急事件的被试都是自愿到实验室来参与实验的。这些人可能都比一般人更愿意帮助别人。为此，拉坦和达利设计了另外一个有趣的实验，希望用另一个情境来观测这一现象。他们找到一个愿意合作的卖酒的商店，该商店同意假装店里发生盗窃事件。当收银员跑到店铺的后面去为一个"顾客"拿啤酒时，该"顾客"（实际上是研究者的同谋）拿了一箱啤酒走出商店大门。这一假盗窃案的上演，总是安排在收银台前正好有一两位真正顾客目睹这一事件的时候。收银员回来后问这一或两个顾客，"嗨，刚才在这里的那个人哪里去了？你看见他离开了吗？"这样，就给顾客一个机会向收银员报告刚才发生的盗窃事件。与实验结果一样，当有另一个旁观者在场的时候，亦即有两个真正顾客在场的时候，向收银员报告盗窃案的行为减少了。

并非仅仅是社会心理学家企图在不同的情境中重验其研究结果。认知心理学家们也开始探索如何将他们的许多实验结果向外推广。例如，吉格瑞恩（Gigerenzer，1984）研究了"频率—效度效应"的可推广性。这个效应是指，一个不熟

悉但看似有理的论断，不管是真是伪，仅仅经过不断地重复，也会增加人们对它的相信程度。这个效应成功地得到了多次验证，但是这些研究都是在实验的情境中，用大学生为被试（并且绝大多数在美国）。于是吉格瑞恩做了一个非实验室情境下、用非大学生作被试的研究。他在德国慕尼黑的许多非大学生的成年人家中进行验证，也发现有"频率—效度效应"，而且其程度与实验室中美国大学生被试所得的几乎相等。

我们将在第十章讨论的"概率推断"原理许多都源于实验室研究，但同时也在现场研究中得到证实。例如，研究者们曾经探研过诸如外科医生、股票经纪人、陪审员、经济学家和赌徒等人士，如何在其特殊的工作场境中做出概率推断（Belsky & Gilovich, 1999; Gilovich, Griffin, & Kahneman, 2002; Kahneman & Tversky, 2000; Wagenaar, 1988）。行为决策理论的原理也已经应用在许多现实生活的情境中，例如，丹佛市警察局用它来决定使用哪种型号的子弹为最佳（Hammond, Harvey, & Hastie, 1992）。亚利桑那州也用它决定是否需要在它的中部修建一座大坝（Hammond et al., 1992）。

实验与非实验结果的证据集中也成为教育心理学领域的突出特点。例如，针对不同课程安排所做的实验研究和现场研究都表明，早期语音教学有助于阅读技巧的习得（Ehri, Nunes, Stahl, & Willows, 2001; Pressley, 2002; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001, 2002; Snow, Burns, & Griffin, 1998; Stanovich, 2000）。

简言之，当前的心理学研究使用了不同种类的实验技术和情境。尽管对许多问题的研究有时候过分侧重使用某一类的技

术，然而与几年前相比，心理学中不同研究方法的频率分布已经比以前更加平均了。

向更精确的方法前进

对于某一问题的研究，我们通常是从相对比较弱的方法向可以做出较强结论的方法转移。例如，我们常常是受到某个非同寻常的有趣个案的激发后，才产生某个特定假设的研究兴趣。正如我们在第四章中讨论的，这反映出个案研究的真正作用所在：为进一步研究某问题提供一些假设，同时激发科学家们用更为严格的方法来研究这些假设。紧接着个案研究之后，研究者多采用相关研究，以确认变量之间是否真正存在联系，而不只是存在于几个特殊个案里的巧合现象。如果相关研究证实了变量之间的关联，研究者就开始尝试采用实验法来操纵相关的变量，借以找到变量之间可能存在的因果关系。于是这就产生了一个从个案研究到相关研究，再到实验操纵变量的研究方法序列。尽管并非每个研究领域都遵循这一逐渐向精确研究方法前进的规律（有时不同类型的研究会同时进行），但这一研究方法序列的确是比较常见的。

在讨论向更精确的研究方法进军之前，必须纠正一些读者可能从第五章的讨论中产生出来的误解，认为相关研究在科学中是无用的。的确，当面临要验证一个因果假设，并且相关法和操纵变量的实验法同时可行时，操纵变量法可能更受青睐。但是，这并不意味着，相关研究对人类知识毫无贡献。首先，许多科学假设都以相关或缺乏相关的形式来表达的，因此相关研究是直接在验证这些假设。第二，尽管相关无法确定

因果关系，但因果关系却确保有相关。也就是说，尽管相关研究无法清楚地证实一个因果假设，但它可以起到排除因果假设的作用。第三，相关研究要比它们看上去更有用一些，这是因为最近发展出来的一些复杂的相关设计可以让研究者做出相当有限的因果推论。在第五章中讨论了偏相关这种复杂相关技术，通过这一技术，我们有可能检验出是否是一个特定的第三变量造成了观察到的一个关系。

我们常用相关研究的最重要原因在于，有时出于道德的考虑，无法对一些变量进行操纵（例如，营养不良或肢体残障）。其他变量，诸如出生顺序、性别、年龄等，则因其无法被操纵，而存在着它们之间本有的相关，涉及它们的科学知识也因此必须建立在相关证据基础上。当然，这一情况并不是只有在心理学领域才有。天文学家们显然无法操纵所有影响其研究物体的变量，然而他们依然能够从中做出结论。

有一个实例可以阐明，在健康心理学中这种研究方法的演化。这个例子就是有关 A 型行为模式与心脏病之间的联系（Curtis & O'Keefe，2002；Friedman & Booth-Kewley，1987；Friedman & Ulmer，1984；Miller，Turner，Tindale，Posavac，& Dugoni，1991；Wright，1988）。最初，A 型行为模式这一概念源于二位心脏病专家的观察，这二位医生从他们一些病人的行为那里发现了一个稳定的模式，这种行为模式包括时间紧迫感，目标不定的敌意，以及对成就的极度追求。于是，一些医生通过对少数个案的观察，提出了 A 型人格这一想法。这些个案研究推出了这个概念，但这还远远不足以作为支持一种假设的有力证据，这个假设就是，一个特定的行为模式是造成心

脏病的原因之一。要证明这一点，我们需要的不仅是少数几个个案研究，它还需要由心脏病专家、生物化学专家和心理学家所组成的研究队伍数十年的努力。

对该问题的研究迅速从简单的病例积累向使用更精确的研究方法的方向不断前进，因为如果只是用个案研究法，这一假设的真伪永远难辨。研究者在朝向用一个比较精确的研究方法前进时，必须进一步发展出A型行为的操作定义。通过大范围流行病学研究，研究者先在A型行为和心脏病之间寻找并找到了相关。接着，这种相关研究工作变得更为深奥复杂。研究者利用复杂相关技术来寻找潜在的第三个变量。由于行为模式与其他传统心脏病风险因素（例如吸烟、肥胖和血液中的胆固醇水平）都有相关，所以很有可能A型行为与心脏病的关系是虚假的。然而，研究结果表明，A型行为可以是一个用来预测心脏病的独立因素。这是因为，当其他可能影响该变量的因素被统计学的技术加以排除之后，A型行为与心脏病之间的关联仍然存在。

最后，研究者利用实验研究来操纵变量，借以探讨两者之间是否可以找到因果关系。有一些研究试图验证是否是一些生理机制影响了两者之间的关系，并采用动物作为实验被试——有些会认为是 “不是真实生活”的研究方法。另一个实验研究用曾经犯过心脏病的人做被试。这些被试被随机分派到两个实验组中去。一组接受辅导，目的是帮助他们戒掉诸如吸烟，过度摄取高脂肪食物等传统心脏病的高危行为。另一组也接受辅导，不过目的是通过一种程序，来帮助其减少A型行为。三年后，发现后一组病人心脏病复发率明显较

低。简言之，一系列证据集中地指向支持A型行为可作为心脏病的重要原因之一这一假设。对这个问题的探研过程可以说是一个很好的例子，用来说明研究是如何逐渐地从有趣的个案研究转化为相关技术，再到更复杂的相关技术，最终达到可操纵变量的实验研究。

从这个例子中学到的最后一个教训就是，科学构念总是在不断地发展之中。这个结论在第三章讨论操作定义时首次提出。最近的研究又倾向于表明，从整体上来看，似乎过分简化了A型行为与心脏病之间的联系。原因在于，只有该构念中的特定成分（特别是对抗性敌意）才与心脏病有关联（Curtis & O' Keefe，2002；Dembroski & Costa，1988）。于是，这个例子告诉我们，科学在前进过程中，如何发掘不断地趋向具体的、精确的关系，以及理论构念是如何被细化的。

对绝望的劝喻

证据集中原则的最后一个意义在于，当对一个问题的最初研究结果之间有点自相矛盾时，我们不应陷入绝望中。科学证据的积累过程就像一个投影仪慢慢地将一张内容未知的幻灯片调好焦距。起初，屏幕上的模糊影像可能代表任何东西。然后，随着一点点地调准焦距，虽然图像仍不能被完全清楚鉴别，但许多原本以为可能是什么东西的假设已经被否决掉了。最后，当焦距调准，就可以非常有信心地做出最终的鉴别。积累证据的过程在早期就如同调焦过程之始。幻灯片的模糊影像就如同互相矛盾的数据，或者可类比为支持许多可能不

同假设的证据。

所以，我们不应该对研究初期所得到的相互矛盾之数据而深陷绝望，认为永远也无法找到真相了。这种情形并非只在心理学中才有。它也发生在许多比较成熟的学科中。实际上，公众不了解的一点是，科学中得到相互矛盾的数据是家常便饭。这些矛盾数据充其量只不过是说明我们对那些数据缺乏充分的理解。它们可能源于随机事件（在第十一章中将对此展开讨论），或者源于实验方法上的细微差异。

许多其他学科在取得内部共识之前都经历过一段不确定的混乱期。格兰德维尔（Gladwell，1996）叙述了有关对脑创伤患者进行适当紧急救治的近期思路发展过程。他讲述了一个幸运的纽约病人得到了世界上最有名的专家，简姆·加哲（Jam Ghajar）医生为他治疗，加哲医生一直致力于改变这一领域的临床医生所持有的一个错误想法。格兰德维尔描绘说，若干年前，当加哲和其他五位研究者在对一些创伤治疗中心进行一项调查研究时，他们发现尽管类固醇已经被反复地证明无助于减少颅内压力（而且会带来潜在危害），然而仍有75%的在昏迷状态的病人是用类固醇来治疗的。他写道：

部分问题在于，在神经外科学这一领域中，难以对治疗程序和方法达成一个可信的、科学的结论……再则，大脑的复杂性和神秘性导致形成一个奖励用直觉来行事的文化，并让每个神经外科医生都坚信自己的经验与其他人的经验同样可靠（p.39）。

几年前，讲到自己同事们的观点停滞不前时，加哲写到，“并不是神经外科专家很懒，而是相关的数据及信息太多，以致

造成混淆”（p.39）。

简言之，就如同心理学的许多领域一样，这一领域也是到处都有人在做研究，但是焦距还没有调好，以及想好要怎么去看这些数据以便能从中找到集中的趋势。于是，1994年，加哲和一些同事参加一系列的会议，试图综合当时所有的证据，借以探看是否可以从这些证据中找到一个集中点。在这个由脑创伤基金会主办的会议上，研究者们就脑创伤处理的14个方面考察了4 000多篇科学论文。脑创伤基金会的执行主席是这样描绘外科医生们开展的工作：“他们所做的工作就是对科学文献中所记载的证据进行论证，一旦有人说，‘这是建立在我自身的经验基础上’时，每个人都嚷道，‘不行，这不算数，根本等于没说。我们想知道的是证据到底在哪里’”（Gladwell，1996，p.40）。这种论证方式最终被证明是有成效的：

参加会议的人发现文献中的确可以找到集中的趋势，因此出版了一本书来展示所有的科学证据，以及脑创伤处理各阶段的最先进治疗法。这本指导手册是神经外科界首次成功地建立的一套标准治疗规范，如果美国大多数创伤中心都能严格按照这些规范去做，他们可能每年可以多挽救上万个病人。（Gladwell，1996，p.40）

尽管由加哲倡导的指导手册已经被用于救死扶伤，有趣的是，当加哲本人谈论这段医学史时，他却特别强调证据集中的价值以及科学知识的公开性（本书第一章中讨论过这一原则）：“人们总是想把它归功于具体的个人，……我猜想这就是

人类本性吧。他们想说：‘这是加哲的规范。他是个了不起的医生。’但事实并非如此。这些规范是建立在当前最好的科学证据之上的。这是每个人都可以拿来用的规范”（Gladwell，1996，p. 40）。

从这个例子可以清楚地看到，并非只有心理学研究结果表现出分散各地的现象。在其他学科里也充满了类似的、因为资料所显示的趋势太过模糊以致难于作出结论的现象。然而，我们也清楚地看到，我们仍然可以找到一些方法，来帮助我们找出一些初步的、暂时的结论。

正如在第七章中所讨论的，从实验结果中发现造成矛盾的原因必将大大促进我们对某个现象的了解。心理学中，多数的“幻灯片”都还十分模糊，因此坚忍不拔就显得双倍地重要，因为当那些困扰了我们几个世纪的、有关人类行为的、复杂而又有趣的问题最终得以解决时，我们也会得到了双倍的快乐。那些无法在模糊的氛围里工作的人通常都不会从事科学工作，因为科学就是充满了不确定性。它所呈现的模糊幻灯片可能破坏了某些人的美感。然而，科学家应有的态度是，对那些目前还很模糊但最终将会变为清晰的影像保留一种强烈的好奇心。这就是布朗诺斯基所描绘的那种“科学是对我们可能知道的东西的一种奉献，虽然在这个过程中，我们有可能犯错误”（Bronowski，1973，p.374）。

小　　结

在本章中，我们看到，为什么科学进步的“突破”模型不适合于心理学，以及为什么“逐渐整合”模型则为心理学提供了一个比较好的、下定论的框架。证据集中原则描述了心理学是如何整合研究结果的：没有任何一个实验可以一锤定音，但每个实验又帮助我们去排除至少若干个其他的可能解释，因而有助于我们追求事情的真相。使用一系列不同的方法，可让心理学家们确信其结论是建立在坚实的实征基础之上的。最后，当研究构念发生变化时，它必须遵循关联原则；亦即，新的理论不仅必须能够解释新的科学数据，而且还必须能够对先前已存在的数据提供一个解释。

（范　照　译）

第九章　别相信有一枪即中的“魔弹”

——多样原因的问题

在第八章里，探讨了操作集中的重要性，以及逐步朝向使用更精确的研究方法的必要性。这两者对于在变量之间建立更精确的、单一的联系非常有用。但是在这一章中，我们将不再只看两个变量间的单一联系，而将焦点放在对另一个重点的思考，那就是，人的行为是取决于多种因素的。

任何一个具体的行为都不是由一个单一的变量引起的，而是由许多不同的因素决定的（Wachs，2000）。当我们得出变量A和行为B之间有因果关系时，并不意味着变量A就是引起行为B的惟一因素。例如，有研究者发现，收看电视和学业成绩之间有相关，但他们不会就此认为看电视是影响学业成绩的惟一因素。道理很简单，学业成绩在一定程度上还受到大量其他因素的影响（例如，家庭环境、学校教育的质量等等）。实际上，相对于这些其他变量，看电视只是影响学业成绩的一个次要因素而已。同样地，收看大量的暴力电视片也不是使儿童表现出攻击行为的惟一原因，它只是众多影响因素中的一个。

和本书中谈到的许多其他原则一样，认识到事物具有多样的成因这一点非常重要。一方面，它提醒我们不要过于依赖

单一的原因解释。因为世界如此复杂，影响行为的因素也相当繁杂。我们虽然可以证明某一变量引起了某一行为，这并不代表已经发现了影响该行为的惟一、甚至是最重要的因素。为了充分解释某种具体行为，研究者必须探讨各种不同的变量对它的影响，并把这些研究结果整合起来，才能完整地描绘出所有与该行为有关的因果关系。

另一方面，虽然说一个变量只是影响某一行为的众多因素之一，因此它只能在很小程度上解释这一行为；但是这并不等于说，这个变量不重要。首先，两者的这一关系可能具有深远的理论意义。其次，这一相关关系可能具有应用价值，尤其当那个影响变量是可以人为地来控制时，如前面提到的电视暴力一例，是可以在电视上少播放一点儿暴力影片的。如果控制了这一个变量，能够使每年的暴力事件降低1%的比率，那我想没有人会认为它是无关紧要的。总之，如果这些负面行为会对社会造成很大影响，那么懂得如何去控制其中的一个成因，哪怕只是很小的一个成因，也已经是相当有用了。罗森塔尔（Rosenthal，1990）举过一个治疗心脏病的例子，在一个实验中，某种治疗方案能将患者存活率提高不到一个百分点；然而，即使这样，这个结果也被认为意义重大，以至于出于职业道德的考虑，实验者不得不提早结束该实验。原因很简单：既然实验治疗结果这么有效，对那些被随机分派在控制组的病人，让他们仍然使用安慰剂，显然很不道德。

交互作用的概念

原因多样化的观点引出了另一个重要概念，那就是交互作用。这个概念在许多方法论的书上都会详细的讨论，所以这里不再赘述，只是提一提：当影响行为的一个因素和另一个因素共同起作用时，它们对该行为会产生与各自单独起作用时截然不同的影响。这就是我们常说的交互作用：一个自变量的影响效果依赖于另外一个自变量的水平。赛蒙斯（Simmons），博格森（Burgeson），卡尔顿－福特（Carlton-Ford）和布莱斯（Blyth）指导的一个实验提供了这样的一个例子（Simmons，Burgeson，Ford & Blyth，1987）。有研究者考察了一组青少年的学业平均成绩，想看看生活事件（如转学、青春期发育、早恋行为、搬家和家庭破裂等）是否会对学业产生影响。他们发现上述生活事件加在一起，是影响学业成绩的关键变量。

另一个例子是努特（Michael Rutter）综述了对与儿童精神疾病相关因素的研究，他归纳道：

最初引人注目的发现是，在实验中，那些被单独分离出来的长期压力来源，各自并没有诱发精神疾病的危险……这些高危因素，当单独去看它们的作用时，都没有与儿童的精神疾病联系在一起；这些儿童得精神疾病的风险也不会比没有家庭压力的儿童高。然而，当任何两种不同来源的压力同时作用时，得病的风险就上升到原来的4倍。若是3种或4种压力来源同时作用时，那么得病的风险就更是增大了好几倍。很

明显地，这些长期压力来源的共同作用远远超过其各自效果的累加。那是因为几种并发压力之间有交互作用，才令其总体效应远远大于个别压力效应的总和（Rutter，1979，p.295）。

要理解交互作用如何发生的逻辑，我们可以像努特所描述的那样，先想像有一个测量风险的量尺，尺上的数字 80 ~ 110 代表低风险，110 ~ 125 代表中等风险，而 125 ~ 150 则代表高风险。假设我们发现儿童在无压力情况下的平均风险指数为 82，在压力因素 A 作用下的平均风险指数为 84，而在压力因素 B 作用下的平均风险指数为 86。那么要同时研究因素 A 和因素 B 两者对儿童的共同影响时，如果发现风险指数达到了 126，也就是说，并发的风险指数远远超过了独立研究单一因素时所预测的结果，就说明了因素 A 和 B 之间存在着交互作用。

发展心理学中也有很多类似于努特所述的例子。邦尼·布瑞特米亚（Bonnie Breitmeyer）和克雷格·拉米（Craig Ramey）研究了两组婴儿，一组是非最佳围产期的婴儿（nonoptimal perinatal），另一组是正常婴儿的（Breitmeyer & Ramey，1986）。在这两类婴儿出生后，再把他们按随机分派成两组——实验组及控制组。然后对实验组进行一个特别的育婴方案，该方案是为了防止出现轻微智力迟缓而设计的。控制组的婴儿则没有得到任何特殊的照顾。当这些孩子长到 4 岁的时候，对他们的认知发展能力进行测试，发现在特别育婴方案下接受照料的儿童中，在围产期不佳的情况下出生的儿童与正常儿童在认知能力上没有明显差异。但是，没有得到特殊照顾的控制组中，那些围产期不佳儿童的表现低于正常儿

童的认知发展水平。该研究中，生理和环境因素的交互作用说明了，一个复杂的行为结果（认知发展）是由多样因素决定的。当在围产期不佳的情况下出生的儿童，又得不到适当的照顾时，就会出现最糟糕的认知发展结果。研究者们总结道：“这个研究结果支持了这样的一个理论架构，即对于那些在社会经济条件低下的家庭中成长的儿童而言，一出生时的生理缺陷会和后天不良的环境因素产生交互作用，从而成为了他们的认知发展的一个累积性危害因素”（p.1151）。

同样地，另一项有关验证抑郁症的“体质 – 压力理论”的研究（Metalsky & Joiner，1992）表明，消极的生活事件与 3 种脆弱的心理因素相结合，会导致非常高的抑郁症发病率。这 3 种因素是归因方式（倾向于把消极事件归因为一些稳定的、整体的因素），对自我进行消极推断的总体倾向，和对任何行为后果进行消极预测的总体倾向（见 Alloy，Abramson，& Francis，1999）。

很多消极的行为和认知后果都是由相似的逻辑所造成。例如，儿童攻击行为是由遗传基因和不良社会环境相互作用所造成的（Pennington & Ozonoff，1996）。同样地，佩迪特（Pettit），贝茨（Bates），道奇（Dodge）和梅斯（Meece）发现，青少年在成长的早期，如果在他们不受管教的情况（例如，下课后或不在家中受父母监管时）下，有大部分时间是与同伴交往，同时在家里也很少受到父母之监管的话，他们会更容易表达出自己的问题（Pettit，Bates，Dodge & Meece，1999）。

正面的结果也可用多样因素及其之间的交互作用来解释。

耐特（Knight），约翰逊（Johnson），卡罗（Carlo）和埃森博格（Eisenberg）在研究 6～9 岁儿童的亲社会行为时，探研了影响儿童助人行为倾向（如：捐款给有需要的儿童）的心理因素（Knight，Johnson，Carlo & Eisenberg，1994）。他们发现一些变量，诸如同情心、情感推理和对金钱的认识等，当单独地去看它们的作用时，并不起眼，它们和亲社会行为之间的相关很低。但是，当将这些变量放在一起来看时，它们对亲社会行为就有很高的预测能力了。例如，具有较强的同情心、较强的情感推理，并对金钱有所认识的儿童，他们捐款的可能性是那些在这些变量上表现较低之儿童的 4 倍。

因此，原因多样化这个概念可能比你最初想到的要复杂得多。我们不仅要追踪并测量出影响问题行为的种种可能因素，还必须研究这些变量放在一起会产生什么作用。

做单一因果解释的诱惑

复杂事件是由多样的原因所决定的，这个观点乍看起来似乎很容易理解。实际上，当问题没有什么太大争议时，这个观点确实很容易掌握和运用；但是，科学工作者的强敌——事先存在的偏见（参见第三章的讨论），却往往使我们忽视了原因多样性这一原则。我们经常听到人们在激烈地争论一些话题，如犯罪的原因、财富的分配、歧视妇女和未成年人、贫困的原因、体罚的作用，以及纳税标准等时，大家的论点似乎总是让人觉得这些问题是非常简单的，而且导致结果的原因只有一个。这些例子进一步反映了尼斯贝特和罗斯

（1980）的说法：“虽然人们通常也承认原因的多样性，但是，显然地，人们在行动上的表现常常和这一观念背道而驰，他们倾向于用单一的因果关系去解释任何事物。在某种意义上说，人们的行为好像是基于一种认为，各种原因是一种“水压式”的关系，或者各种可能原因之间存在着像是在玩“要不全输—要不全赢”（俗称“零—和”）游戏那样，是一个相互竞争的关系”（p.128）。

“零—和”游戏——一个人的所赢的正是游戏中其他对手所输的——常常是我们在讨论那些容易激动的问题时所采用的思维方法。在情绪高涨的影响下，通常会忘掉原因多样性这个原则。让我们来看看两个敌对的政党是如何讨论社会犯罪问题的，主张自由开放的人会认为那些社会经济地位低下的人之所以会犯罪，是因为他们本身就是恶劣社会环境（如失业、破陋的住所、缺乏教育和对未来丧失希望等）的受害者。而比较保守的人会争辩说，大部分的穷人也都没有犯罪啊！所以社会经济条件并不是主要原因。因此他们认为，个人的价值观和人格特征才是决定犯罪行为的真正原因。在这场争论中，双方都没有谈到，造成犯罪行为的原因既有个体的，也有环境的原因。

政治评论家理查德·科恩（Richard Cohen）也曾在文章中谈到通常是如何根据先验偏见，来做“单一原因”的解释的。他引用了衣阿华州一位63岁的农场主人的例子，这位农场主在该州农业经济进入萧条期时，因为背负上了沉重的债务，即将被银行没收他的农场。在感到还债无门的绝望中，他开枪打死了他借钱的那家银行的经理，随后在枪杀了他的妻

子之后自杀了。在讨论这一事件时，无论是他的邻居还是传统传媒的看法都认为，这个农场主是因为遭受了巨大的财务困难才会“崩溃”的。媒体的报道都对他报以极大的同情。科恩写道：“那位农场主被描绘成是一位勤劳的企业家，他顽强地与大自然、银行及像是在芝加哥农业品交易市场的拍卖员那样的人，进行搏斗。他是一个诚实的、节俭的，以及是大多数美国人优点的化身——他是不靠他人的个体户、自食其力，而且崇敬上帝”（1985，p.11）。

但是，令科恩质疑的是，如果这个人真的只是因为财务困难就去杀人和自杀的话，那么我们是否可以用同样的单一原因理论，去解释那些发生在贫民窟的杀人事件？“如果对于受困的农场主而言，这是其崩溃的原因的话，那么为什么发生在贫民区的情况就不同了呢？为什么当我们提出贫穷、恶劣环境、三流学校和野蛮行径是造成犯罪的原因时，会遭到如此多的责难？”（p.11）。在贫民区发生的同样事件会被视为是个人的失控及责任，而且应当受到严厉的处罚，而如果当它发生在衣阿华州的农场主人身上时，则被认为他是一个“环境的受害者”，公众也因此对他表示同情。科恩进一步指出了，在做单一性原因解释时，所产生的一个荒谬现象：在利用单一解释来巩固已有的偏见。他认为如果从一开始就认识到以下所述的例子中发生农场及贫民窟的杀人事件都可能是由多样因素决定的，那么也许就可以避免犯下这一荒谬的错误。不论是爱荷华州的农场主杀人，还是贫民窟的人之杀人行为都是受其个人的心理－生理特性和环境压力共同影响的。没有一种单一的原因可以单独解释犯罪事件。犯罪行为是受许

多因素影响的，包括环境因素，也包括个性因素。

在讨论复杂的经济形势时也存在相类似的例子。一个曾引起了长达几十年争论的、极具社会政策意义的重要话题是：为什么美国的贫富差距继续不断地扩大（Beatty，1996；Bronfenbrenner，McClelland，Wethington，Moen，& Ceci，1996； Frank，1999； Frank & Cook，1995）。正如本书在第六章中所提到的“聪明汉斯”的例子一样，这一事实不是争论的焦点，而对这一事实的解释才是。从1979年至今，美国全体男工的真正收入（即排除通货膨胀之后）下降幅度超过了10%（Cassidy，1995； Mishel，1995）。当然，也有个别小群体干得很出色，收入有所增加。自1979年以来，持有研究生文凭的职工群体的真正收入增长了8.8%，但是对那些只受过“普通大学教育”的大多数职工来说，他们的真正收入却下降了15.9%（Beatty，1996）。在美国家庭中占20%的中产阶级家庭的平均收入（同样是排除了通货膨胀之后），从1973年到1993年间大体上保持一致，这主要是因为在那段时期，大量家庭妇女涌入劳动力市场，令双收入的家庭增加了。妇女融入劳动力市场这一趋势，其实主要是因为受到工资水平停滞不前或不断下降，而不是纳税增加的影响，靠丈夫一个人的单薪已经无法维持生计所造成的。因为根据收入而定的税率从1979年（31.1%）到1993年（30.9%），一直是保持在非常稳定的水平之上。

与中产阶级和贫困阶级的收入之停滞不前甚至不断下降相比，恰恰是在1973~1995年这一时期里，许多有钱人的财富正在暴长（Frank，1999）。人口总数里收入最高的那1%的

人，从1977到1990年的十余年间，收入增长了74%（这个数值也是在排除了通货膨胀之后的数据，Slemrod & Bakija，1996）；与此同时，美国中产阶级的收入只增长了3%；而在人口总数中占20%的处于社会底层的人，他们的收入却下降了13%。1977年，社会上最富的20%所赚的钱是那些最穷的20%所挣得的4倍。而到了1991年，这一比例增加到7倍了（Frank & Cook，1995）。1973年，最富的那20%的人的收入占美国国民收入总额的43.6%，而在1993年，该比例已经达到了48.2%（Cassidy，1995）。虽然这些百分比的差异看来并不是很大，但是它却带来了深远的社会影响。正如卡西德（Cassidy，1995）指出的：它意味着“在每年的收入中，由中产阶级流入富人阶级之口袋的钱大约有2 750亿美元，平均每个中产阶级的家庭要为这一趋势付出4 500美元”（p.118）。这已足够让大多数美国人的生活发生真正的变化。举个例子来说，这将导致大部分国家公立大学的学费收入会大大减少，因为会有比较少的家庭有经济能力供得起他们的子女上大学。

财富从市民中的一个阶级大规模地转移到另一个阶级手中，这个现象的社会影响引发了一场极富争议的、有关其背后原因的政治辩论。这场争辩最引人注目之处就是，这些争论者都只关注单一的原因。争辩中的每一方都只以某一个原因为立论基础，然后千方百计地攻击所有认为其他因素是原因的观点。事实上，计量经济学研究（Beatty，1996； Cassidy，1995； Frank & Cook，1995； Mishel，1995）已经重点探研了4个变量（研究者所提出的变量的数目远不只4个，但这

4 个得到最广泛关注及研究的）。其中一个因素是科技。例如，论据之一是，懂得使用计算机的人提高了生产效率，他们的工资也因此而增加了。与之相应的是，计算机还替代了很多没有特殊技能的工人（如邮件分类员、银行出纳员等等）的工作，因而就降低了他们的工资收入（Cassidy，1995）。争论的第二个因素是新移民的不断涌入美国，而这些人多是缺乏技能的工人，他们使得已经很低的工资水平继续下滑，因为没有技能的劳动岗位本来已经供大于求了。第三个原因是全球化，它助长了进一步的收入不均，是因为公司可以搬到国外，只需付很低的工资，就可以雇佣一些没有技术以及稍有技能（但会逐渐地变成有技能）的工人，而这更加重了本国低技能劳动力的过剩（Cassidy，1995）。第四个原因是工会联盟作用，这些联盟的影响力正在不断地下降，而大企业的控制力却在逐步增长。论据就是，在 20 世纪 90 年代，劳工的罢工事例减少了，而资本的罢工事例（即企业放弃了某一区域为其生产基地，因为在其他区域生产可以取得更高的资金回报率）却增加了，从而提升了资本的价值而降低了劳动力的价值。

经济学研究这 4 个变量的时候到底发现了什么呢？哈！你猜对了。所有这四个因素共同造成了社会的正在崛起的不平等。这个例子也阐明了先前所提到的、交互作用的概念。卡西德（1995）在文章中指出，“某些因素可能有交互作用并相互强化。随着全球竞争的日益激烈，企业经理们将更倾向于不理工会的加薪要挟，将资本投入可以取代人的计算机技术。同样地，公司迁移的威胁和外来廉价劳工的增多都会导

致工会力量的进一步削弱”（p.122）。

实际上，和经济学的问题一样，心理学所研究的复杂问题也都是受到多样因素的影响的。举个学习障碍的例子来说明之，这个问题已经被教育心理学家、认知心理学家和发展心理学家广泛地研究过。结果发现，脑部的病变与学习障碍有关（Galaburda，1994；Hynd，Clinton，& Hiemenz，1999）。也有研究发现，学习障碍有一部分与遗传有关（Cardon et al.，1994；Olson，1999）。这两个研究结果看起来好像可以让我们做出一个结论：学习障碍仅仅是属于生理——脑的问题。如果很快地就下这一结论，那就错了（Sternberg & Spear-Swerling，1999）。因为进一步的研究发现，造成学习障碍的部分原因是，早期学校教育中缺少某一些教导体验（Pressley，2002），以及家庭环境不好（Senechal & LeFevre，2002；Snow，Burns，& Griffin，1998）。学习障碍因此不是由单一原因所引起的；相反地，它是生理与环境因素交互作用的结果。

小　结

本章内容虽然简单，但却是非常重要的。考察行为的原因时，要依照多样性的原则来思考。不要陷入类似于某一具体行为只由某一特殊原因所造成的这样一个误区。大部分复杂的行为都是由多样因素来决定的。各种各样的因素共同起作用才会引发某一行为。有时多个因素联合在一起时会产生交互作用。也就是说，变量合在一起的整体效应，远远大于当我们

试图只研究他们各自在单一变量因果研究中所找到的单一效应之总和。

（郭妙燕　译）

第十章 人类认知的死穴
——概率推理

问：男人比女人高，对吗？
答："对。"

问：所有男人比所有女人都高，对吗？
答："不对。"

答对了啦！信不信由你，在本章中，我们将要花一些篇幅来讨论，从你刚才的回答中看出，你已经知道答案的问题。但是，先别因此就跳过这一章！因为在下面当我们去解释一个看起来非常简单的原则时，你会发现一些新东西。

你在回答第一个问题时，你给的是一个肯定的答案，这是因为你没有把"男人比女人高"这句话理解成第二个句子所说的"所有的男人比所有的女人高"。你把第一句问话正确地理解为"男人有比女人高的趋势"的意思，因为每一个人都知道，不是所有的男人都比所有的女人高。你理解到那句问话反映了一个概率趋势，而不是一个在任何情境中都适用的事实。概率趋势（probabilistic trend）是指有较大的可能性的意思，但是，这并不是说在所有情况下都必然是这样的。也就是说，性别和身高的关系要用可能性和概率的词语，而不是用必然的

字眼来描述。在自然界中很多关系的本质也是概率性的，例如：接近赤道的地方天气比较热；每家的孩子数目不超过8个；地球上大部分地方的昆虫数量比人的数量多。这些都是可以用在统计学上的概念表达出来的趋势，但是它们当中的每一句话都不是绝对的，仍然可能会有例外。因为它们是概率的趋势或定律，而不是在每一个别情况下都成立的关系。

事实上，心理科学所揭示的所有的事实和关系都是用概率来阐述的。在这一点上，心理学也不是惟一的学科，在其他学科里，很多定律和关系也是用概率，而不是用必然性来阐述的。例如，人口遗传学的所有子学科都是基于概率关系；物理学家告诉我们，原子中电子的发电分布也是通过概率函数来描述的。

的确，心理学中所揭示的大部分概率趋势都比其他学科要弱一些。在心理学中，各种行为关系都是以概率形式加以描绘的，然而这一事实并没有使得它与其他学科之间产生天壤之别。布朗诺斯基指出（1978a），随着科学不断地开拓出新的研究领域，越来越多的科学定律都将采用概率术语加以描述，对于这一点，许多人不能接受：

如果我说，一周中，经过前6天的好天气之后，星期天总是会下雨，这就会被认为是一个定律，因而受到重视。但是，如果我说，一周中，经过前6天的好天气之后，星期天下雨的可能性比不下雨的可能性要大，这就是一个不太令人满意的说法，并且人们会理所当然的认为，我没有真正把一个定律找到。因为定律与通常人们习惯于想从科学中得到那种决定性的“是”或“否”的结论是不谋而合的。甚至如果

我说，一周中，经过前6天好天气之后，10个星期天里，有7个会下雨，你可能会把它当作为一个统计数字来接受，但是它还是不能让你满意，因为它不是一个定律。它看上去就是不如一个定律那样有威力。然而，这纯粹是一种偏见……我在这里所解释的概率的理念你不难理解，，但是它很新且不为人熟知。我们过去没有接触过它……虽然我们生活在充满了“有时”和“也许”的世界里，而我们的愿望却是要生活在“总是”和“一定”的世界里。……然而我认为这个困难仅仅是一个习惯。一旦我们愿意，一旦需要去熟悉这些新的观念，我们就会对它们逐渐熟悉。我们现在已经到了非要去熟悉这些概率理念的时刻了（pp.81～82，94～95）。

在这一章里，我们将尽力使你能更好地把握“有时和也许的世界”，因为要了解心理学，必须熟悉本章所要讨论的主题：概率推理。

“张三李四”的统计学

众所周知，医学科学中很多的结论是用概率趋势来描述的，而不是用必然性来描述的。吸烟会导致肺癌和种种其他健康问题。大量医学证据都证明了这个事实。那么每个吸烟的人都会得肺癌，而每个不吸烟的人就一定不会得肺癌吗？大家都知道并不是这样的。这种关系就是概率关系。吸烟在很大程度上增加了患肺癌的几率，但并不是必然导致肺癌。医学科学可以很有信心地告诉我们，吸烟的人与不吸烟的人相比

较，会有更多的人死于肺癌。但是这一信息不能告诉我们是哪一些人会死亡，这种关系就是概率关系；它并不是在每一个情况下都适用。我们都知道这一点，不是吗？恐怕不一定吧？我们经常看到下面这样的场景：一个不吸烟的人引证吸烟导致肺癌的统计数据，企图说服一个瘾君子戒烟，所得到的结果仅仅是对方的反唇相讥：“喂，走远点儿！你看那个店铺里的张三，他从16岁开始，每天要吸三包骆驼牌烈烟！现在他已经81岁了，看上去还非常健壮！”任何人听了这段话就只能得出一个结论，那就是这一个案例已经推翻了吸烟与肺癌的关系。

令人吃惊和沮丧的是，这种反驳手段屡试不爽。每当一个单一的个案被用来证明概率趋势无效时，很多人都常常点头表示赞同。他们的赞同反映出他们没有正确理解统计定律的本质。如果人们认为单一个例子就可以使一个定律失效，他们一定认为这个定律应该在任何情况下都适用。简言之，他们错误理解了概率定律的性质。既使是最好的趋势也会有少数的“张三李四”与之背道而驰。就拿吸烟的例子来说，活到85岁的人中只有5%是吸烟者（University of California，Berkeley，1991）。或者从另一方面来看，活到85岁的人中有95%属于从不吸烟者，或在一段时期内吸烟但最终停止者。连续从未间断地吸烟者会明显减少寿命（University of California，Berkeley，1991），然而，当然确实有少数长期吸烟者能活到85岁。

借用心理学家尼斯拜特和罗斯（1980）的术语，我们将把类似“张三”的故事称作“张三李四”统计学原理的一

个应用范例：由于一个人知道几个“张三李四”与某一个成熟的统计学趋势相违背，这个趋势就会被人怀疑。例如，我们经常听到类似的话——“你说在服务业中就业机会在扩大；而在重工业中正在缩小？这不对，我就知道有一个“李四”上周四就在一个钢铁厂找到了一份工作”；“你说与30年前相比，家里的孩子少了？你在胡说！隔壁的年轻夫妇已经有了3个小孩，但他们还不到30岁”；“你说通常孩子都会倾向信仰他们父母的宗教信仰？但据我所知，我的一个同事的孩子就在前几天改信了另一个宗教。”

当我们面对和过去持有的观念相矛盾，同时又是强有力的统计证据时，无所不在的“张三李四”总是会立刻跳出来否定这些证据。因此，我们可以说，实际上人们比我们想像地知道的更多，他们只不过顺手把“张三李四”当作一个工具，来把与他们观念相违背的事实给否决掉而已。然而，研究人类决策和推理的心理学家们的研究结果表明，人们之所以使用“张三李四”，不仅仅只是因为它在争论时是一个很有用的战略手段。反而，这一错误的争论模式之所以被应用得如此频繁，主要在于人类不知道如何处理概率信息。决策心理学的最新研究表明，概率推理可能正是人类认知的死穴。

概率推理以及人们对心理学的误解

在科学、技术和人文事物的许多领域中，都涉及到概率思维。所以，我们也没有什么特别的理由认为这种思维对理解心理学比对理解其他学科更重要。然而，由于人们在运用概率信

息时所存在的问题，心理学的研究结果常常被误解。我们都理解"男人比女人高"是一个概率趋势的陈述，所以并不会因为有例外的情况，就认为这一陈述（一个男人比一个女人矮）是错的。很多人也能以同样的方式来理解"吸烟可以导致肺癌"的陈述，尽管对那些不想相信生活习惯会致他们于死地的瘾君子们来说，"张三"可能还是具有说服力的。然而，与这些例子的陈述相似的、但是却是关于行为趋势的概率陈述，却会引起很多人的怀疑，而且常常在一开始，就被他们用一个"张三或李四"的个案给打发掉了。很多心理学教师在讨论某些行为之间关系的证据时，都往往得到同一个反应。例如，教师可以呈现如下的事实，儿童的学业成绩和家庭的社会经济地位及父母的教育水平相关。但这个事实常常会遭到至少一个同学的反对，他会说，他有一个朋友，是一个全国知名的学者，但是他的父亲只是中学毕业。甚至那些理解吸烟－肺癌例子的人，对这一问题也变得摇摆不定了。

人们从没想到过要用"张三李四"的论据来反驳医学和物理上的发现，却习惯于用他们来驳斥心理学的研究结果。大多数人能理解医学科学提出的治疗、理论及事实是概率性的。例如，他们理解一种药对一群病人来说，并不是对他们各个都有疗效，而且医学也经常不能事先告诉我们，对哪些病人会有疗效。通常可以说的是，100个病人接受治疗A，100个病人不接受任何治疗，在一段时期之后，接受治疗A的这100个病人，作为一个总体来讲，会比不接受A的100个病人的病情好一些。没有人因为这是一个概率陈述，因而不是在所有情况下都适用，就怀疑这一治疗的价值。许多心理学的研究结果和心理治

疗的疗效，也存在类似的情况。然而，一旦心理学研究结果和心理治疗疗效不能在所有情况下都适用时，就常常会引发人们对心理学的进展产生极大的失望和轻蔑。一旦面对心理学的议题，人们常常忘记了一个最基本的原则，那就是知识不需要100%确定后才是有用的——即便某些知识不能预测个体的具体情况，但是如果能对一些群体的总趋势有预测能力，也是非常有益的。基于群体的特征所做的结果预测常常被称为是总体统计数字（aggregate）或精算预测（actuarial prediction）（在下一章详细地讨论精算预测这一概念）。

由于这些原因，全面地理解概率推理对理解心理学非常关键。在这里，我们也看到了一个激起人们深思的讽刺现象，心理学很可能是人们不能作统计思维的最大受害者，然而心理学家却是对人类概率推理能力进行最多研究的人。

有关概率推理的心理学研究

过去的20年里，普林斯顿大学的卡尼曼（Daniel Kahneman，2002年诺贝尔奖金得主，见MacCoun，2002），密歇根大学的尼斯拜特以及已故的特维斯基等心理学家的研究彻底改变了我们对人类推理能力的认识。他们在研究期间发现，很多人头脑里不具备概率推理的基本定律，或者即使有也不够用。正如学者经常指出的，这些基本定律在人们头脑里没有发展好是不足为奇的。作为数学的一个分支，统计学是最近才发展出来的（Hacking，1975）。反而是运用几率的游戏，要比概率定律早出现几个世纪。这又是一个例子说

明：个人的经验不足以让人们获得对世界的基本理解（参看第七章）。成千上万的赌徒以及他们的个人经验，并不足以让我们发现几率游戏的本质。反而是通过对概率定律的正规研究，才让我们找到了人们一直以来所玩的几率游戏背后的运作方式。

当一个社会变得越复杂时，我们就越需要概率思维。如果一个普通市民想要对其生活的社会有一个基本的理解，那么，他在最低的限度上，必须具备可以运用统计思维的最基本能力。

你可能经常会有以下的一些问题:“为什么他们要提高我的保险费？”“为什么张三的保费比李四高，是社会保险局搞错了么？”“我们这一州的彩票背后有黑幕？”“犯罪事件到底是在增加还是在减少？”“为什么医生要安排这些检查？”“为什么用非欧洲人可以用一些很珍奇的药，而美国人就不让用？”“在相似的工作岗位上，女人赚的真的比男人少吗？”“国际贸易真的减少了美国人的就业机会，并降低了他们的薪水吗？”“日本的教育成就要比我们好吗？”“加拿大的卫生保健真是比美国好，并且价格低廉吗？”这些问题都问得很好，都是关系我们社会及其如何运作的、具体而且实际的问题。要理解每一个问题，我们就必须以运用统计思维。

显然地，本书由于篇幅的限制，不能全面覆盖统计思维的所有讨论。然而，我们将简要地讨论某些关于概率推理的较为普遍的误区。学习概率思维技巧的最好方法就是，去认识并警觉人们在用统计推理时最常犯的错误是什么。况且，

这些错误中，有许多可以帮助我们进一步看到心理学研究结果及理论的重要性。

没有充分利用概率信息

在心理学领域中，有一个发现是已经得到了多次重复验证的，那就是，有关一个具体事件的信息往往可以完全击败比较抽象的概率信息（在第四章中讨论的“鲜活性”问题）。忽视概率信息的例子比比皆是，而且并不仅仅局限于科学知识不丰富的外行人。卡斯塞尔斯（Casscells），谢诺博格（Schoenberger）和格瑞博维斯（Graboys）在4所哈佛医学院的教学医院中进行了一个研究，他们问了20位医学专业的学生，20位内科主治医师和20位办公室的工作人员下面一系列问题：“如果在每1 000人中有1个人携带艾滋病的病毒（HIV），再假设有一种检查可以百分百地诊断出真的携带该病毒的人；最后，假设这个检查有5%的可能性，把没有携带者说成是有。也就是说，这项检查在没有携带HIV的人中，也会错误地检测出有5%的人是携带病毒者（Casscells，Schoenberger & Graboys，1978）。假设我们随便找一个人来进行这项检查，得到了呈阳性反应，亦即此人为HIV携带者。假定我们不知道这个人的患病史，那么他真的是HIV携带者的概率是多少呢？

一般人会回答95%，正确的答案是接近2%。医生们对一个诊断为阳性的病例到底是不是真的是确诊之统计推断往往过分高估，这是因为他们一方面过分重视个案信息，另一方面又过分忽视基本比率信息，从而过高地估计了阳性测验结果所

真正代表的患病概率。少许的逻辑推理可以帮助说明基本比率对概率思维的重要作用。1000个人当中只有1人会确诊是HIV呈阳性反应。如果另外999人（不患有这种病）也进行了这样的检查，由于这一检查有5%的虚报率，他们当中将有接近50人（999乘以.05）被检查出携带这种病毒。这样一来，呈阳性反应的人就会是51个。因为在这51个人当中，只有1人真正是HIV呈阳性反应者，此人确诊得病的概率其实只接近2%。简而言之，这一假想题的基本比率就是大多数人没有携带这种病毒。这个事实和相当大比例的虚报率放在一起考虑就能算出，在绝对数量上，大部分呈阳性反应的人并不携带这种病毒。

尽管参与卡斯塞尔斯等人之研究的医生，很快地就意识到了以上概率逻辑的正确性，但他们最初的直觉反应却是忽视了基本比率，并过分看重临床检测的证据。简单地说，事实上医生们早知答案不会是那么高（95%）的百分比，但却从资料中得出了错误的结论。心理学家把这类问题称为认知错觉（*Cognitive Illusions*）（参见Kahneman和Frederick，2002）。在认知错觉中，即使人们知道正确答案，他们也会由于问题的问法而作出错误的结论。

在这里，我们所提到的所有例子都是认知错觉，因为它们都是落入人类推理的一个陷阱：过分看重个别事件所提供的证据，而忽视了统计学的信息。对大多数人来讲，个案证据（实验室的研究结果）好像是抓得住的、具体的，而概率证据则好像是抓不住、不确定的。当然，这种理解是错误的，因为个案证据本身一定是概率性的。一个临床检验是会在一定的概

率范围内误诊一个人有某种疾病。前面举的例子就是有两种概率在同时作用——对个案证据做出正确或错误诊断的概率（即95% 或 5%）和过去经验所提供的先验概率（prior probability）（即 0.1%）——要想做出正确的判断必须将二者结合起来。结合这些概率的方法有的是正确的，也有的是错误的，特别是当个案证据给人一种错觉它是比较抓得住及具体的时（请回忆在第四章所讨论的鲜活性问题）——人们往往就会以错误的方式来结合信息了。这个具体的概率推理失败，说不定正是阻碍了心理学知识应用的祸首，因为心理学的知识正是经常用概率的形式来表述人类各种行为之间的关系。

科普作家科尔（1998）要我们想像下面两种情况。一种是用吸烟的死亡率来劝人不要吸烟，这是一种最常见的劝人方式。第二种方法则更为生动一些，让吸烟者想像在每 18 250 包烟中有一包是与众不同的——它里面装满了炸药，当吸烟者打开它时就会被炸死。我们绝对知道哪一个更具有效果——然而他们却是表达了同样的一个事实。

最后，艾滋病的例子是体现概率思维在美国社会之重要性的一个很好的例证。有些人可能认为这是一个虚构的故事，但是，很多年来，在某些职业和政府部门中确实强制实施了艾滋病检查，而且这一问题已经引起了激烈的争论，可悲的是，这一强制检查恐怕会一直继续成为社会的一个问题。正如数学教授林恩 · 斯廷（Lynn Steen）所辩称的：

对强制性艾滋病检查的持续争论，提供了一个很好的例子，说明了隐藏在这些争论之下的，一些有关数量的问题……检查总会有数量很少的（或许 2%）的错误……因此公

众以此来推断这一检查的正确率是98%。但是，因为在普通人中，出现艾滋病的病例，少于检查中误诊的数量，所以对随机取样的市民进行任何大规模的检查，如果检查出一个人有艾滋病，他是被误诊的可能性，其实是远远大于他真的是携病毒者（Steen，1990，p.218）。

没有充分利用样本大小信息

请大家考虑下面两个由特维斯基和卡尼曼（1974）提出的问题：

1. 一个小镇里有大小两所医院。在大医院里每天大约有45个婴儿出生，在小医院里每天大约有15个婴儿出生。正如大家所知道的，大约有50%的婴儿是男孩。当然，真正的百分比每天都不一样，有时候高于50%，有时候低于50%。在一年的时期内，每一所医院都记录了出生的男婴比例高于60%的天数。你认为哪一个医院记录的天数多？

   a. 大医院

   b. 小医院

   c. 基本一样

2. 假设一个缸里装满了球，其中有2/3是一种颜色，1/3是另一种颜色。一个人从缸里拿出5个球发现有4个是红色的，1个是白色的。另一个人从里面拿出20个球，发现有12个是红色的，8个是白色的。哪一个人更自信地认为这个缸里有2/3的球是红色的，有1/3的球是白色的，而不是有1/3的球是红色的有2/3的球是白色

的？这两个人会给出什么样的概率来说明这两种说法的正确程度呢？

回答第一个问题时，大多数人回答“基本一样”。剩下没有如此回答的人们，则一半选择大医院和一半选择小医院。但正确的答案是小医院，所以接近75%的被试都给出了错误的答案。得到这些错误答案是由于人们没有认识到，样本的大小在这个问题中的重要性。当其他的因素保持不变时，较大的样本总是能够更精确地估计出样本母体的真正数值。也就是说，在任何一个指定的日子，较大的医院由于有较大的样本，男婴出生的概率倾向于接近50%。相反，小的样本总是倾向距离样本母体的真正数值比较远。因此，小医院将会有更多的天数记录了与样本母体的真正数值相矛盾的男婴出生的比率（60%的男婴，40%的男婴，80%的男婴，等等）。

在回答第二问题时，大多数人认为5个球的样本提供了更令人信服的证据，可以证明这个缸里的球大多数是红色。事实上，利用概率思维，则恰恰得到相反的结果。在5个球的样本中，在缸里真有2/3为红球的情况下，抓出4个红的、1个白的之概率与在真有1/3为红球的情况下，抓出同样比例的球出来的概率相比是8∶1。而在20个球的样本中，在缸里真有2/3为红球的情况下，抓到12红球、8个白球的概率与在真有1/3为红球的情况下，抓出同样比例的球出来的概率相比是16∶1。尽管在5个球的样本中，抓出的红球之的比例较高（80%），这并不能抵消另一个取样大小为其4倍的样本，在缸中的比例进行推断时有较大的可信度，这一事实。然

而，大部分被试在这个题目作出判断时，多因为在5个球的样本中，红球有较高的比例，而没有充分考虑到20个球的样本有较大的可信度这一点。

认识到样本大小对信息可信度的作用，是对证据进行评估时必须注意的基本原则。这一基本原则，固然经常被应用于许多不同的领域之中，但是其对评估行为科学的研究结果尤为重要。不管我们是否意识到，我们对大一点的群体的确都会持有一些泛化的想法（译者注:例如，“美国人都是不合群的”）。然而，我们最坚定的一些想法是建立在那么脆弱的事实基础上，倒真是我们很少察觉到的。把对几个邻居和几个同事的观察，以及在电视新闻上看到的一些趣闻轶事放在一起，我们就已经觉得自己有资格对人性，或者“美国人”评头论足。这里只是指所用样本的大小，还完全没有谈及样本是否具有代表性的问题。那当然又完全是另外一个问题了。

赌徒的谬误

赌徒的谬误是指人们认为过去发生的和将要发生的两个相互独立的事物之间存在着某种联系的倾向。当一件事情发生了之后，其对另一件事情发生的可能性不发生任何影响，那么，这两个事件可以说是相互独立的。大多数装置有随机设备的赌博游戏就有这个特点。例如，每一次在大轮赌盘上出现的数字，并不依赖于上一次出现的数字。在大轮赌盘上有一半的数字是红色的，另一半是黑色的（为了简单起见，我们忽略绿色的零和双零这两个数字），这样，旋转后得到黑色和红色的数字的概率是一样的，所以每一次旋转后都有可能

得到红色的数字。然而连续5、6次都得到红色的数字后，很多赌徒都迅速的将筹码转而押到黑色那一边，认为下次得到黑色的数字之可能性较高。这就是赌徒的谬误：当事件之间原本是相互独立，但却被认为前一个事件的结果影响了后一个事件结果发生的概率。赌徒们的这一想法是错误的。大轮赌盘对先前发生的事情是没有记忆的。即使一连串有15个红色的数字出现后，下一次旋转得到红色数字的概率仍是50%。

你可以用掷硬币来证明在公众的心目中也存在类似的谬论。如果你问一群人，连续5次掷硬币得到的都是人头朝上，下一次得到人头朝上的概率是多少。一部分人会说不可能再出现人头朝上。这又是赌徒的谬误。每次掷硬币都是独立的事件。连续5次人头朝上后，硬币仍有两个面，所以在掷硬币时每一面都有相等的概率会朝上。

不仅仅是没有经验的赌徒或是初学者才会犯赌徒的谬误。过去研究发现，即使是每星期玩赌博游戏超过20小时的资深赌徒，仍然会表现出赌徒的谬误（Wagenaar，1988）。而且，我们要认识到这一谬论不仅仅局限于赌博游戏，它还存在于任何概率起着重要作用的领域。换句话说，它几乎发生在每一件事情上。婴儿的基因构成就是一个例子。心理学家、医生和婚姻顾问常常看到已有两个女孩的夫妇，在计划要生第3个孩子时说，因为“我们想要1个男孩，这一回一定是个男孩”。这就是赌徒的谬误，在生了两个女孩之后生男孩的概率（接近50%）和生第一个时是完全一样。生了两个女孩之后不会增加第三个孩子是男孩的概率。

赌徒的谬误存在于任何一个有几率成分的地方，例如在体

育运动和股票市场。一些心理家（Gilovich，Vallone & Tversky，1985）已经研究了在篮球运动中对“连投连中”或“很烫手”的迷信，这一迷信是指，相信某一个投篮手能够变得“很烫手”，并且接下来会连续投中，（“球传给他，他现在炙手可热”）。研究者证实在篮球运动员及球迷中，连投连中的想法非常强烈。例如，在一个问卷调查中，91% 的篮球迷认为刚投中两球或 3 球的球员，与刚有两次或 3 次失误的球员相比，在下一次投篮时会有较高的投中概率；84% 的球迷认为，把球传给刚刚连续投中两球或 3 球的球员是重要的。当请球迷估计，假设一个球员在场地上有 50% 的投中率，那么在他投中一球后，再投中的百分比是多少？一次没投中后，再投时投中的百分比又是多少？结果，球迷们估计前者是 61%，后者是 42%。研究者调查了费城 76 人篮球队的队员，结果发现大多数（但不是全部）球员对连投连中所持有的强烈信念与球迷们几乎一样多（见 Gilovich 等，1985）。

可是为什么我们要在赌徒的谬误之大标题下讨论连投连中呢？因为跟本没有连投连中这回事！基诺维奇等人研究了费城 76人队和波斯顿塞尔特斯队在1980 ~ 1981年赛季中投篮命中的统计数据（Gilovich et al.，1985）。在这一赛季期间，球员们的投篮并没有出现前后相互依赖的现象。现在让我们用非技术的语言来看看，这代表什么意思。

赌徒的谬误是把独立的事件加以联系，也就是认为毫无关系的事件之间存在着相互依赖的联系。在统计学上，连投连中是指连续投中两或 3 球后，投篮的命中率高于连续两或 3 次

失误的投篮命中率的假设。基诺维奇等人（1985）计算了这个概率，发现没有任何证据支持这个假设，例如，埃维（Julius Erving，在费城76人队中投篮最多的球员）的资料表明他在连续3次投中后，投篮的命中率为.48，而连续3次没投中后投篮的命中率为.52；在连续两次投中后，投篮的命中率为.52，而连续两次没中后，投篮的命中率为.51；在1次投中后，投篮的命中率为.53，在1次没投中后，投篮的命中率为.51。简单地说，无论前几次投篮的情况如何，埃维的命中率都是接近.50 ——根本没有连投连中这回事。

其他球员的资料也非常类似。莱昂内尔·霍林斯（Lionel Hollins）连续两次投中后投篮命中率是.46，连续两次没中后，投篮命中率是.49。他投中1次后，投篮的命中率是.46，和1次没投中后投篮的命中率是完全一样的。这说明，不管霍林斯前几次的投篮的结果如何，他投篮的命中率总是接近47%。波斯顿塞尔特斯队的罚球资料也说明了同样的情况。例如，拉里·伯德（Larry Bird）在投中1次罚球后下一次罚球投中的概率是88%，而1次罚球没中后，下一次罚球投中的概率是91%。纳特·阿奇博尔德（Nate Archibald）在投中1次罚球后，下次罚球投中的概率是83%，而1次没投中后，下次罚球投中的概率是82%。由此可见，在罚球中也没有连投连中。相信球员可以变得"很烫手"的想法确实是赌徒的谬误的一个例子，也就是说，相信事实上独立的、毫无关系的事件存在着联系。

有趣的是，赌徒的谬误看起来又一次验证了在第六章讨论的"直觉物理学"时所得出来的结论——仅只是凭借经验是无

法告诉人们世界的真相的。基诺维奇等人（1985）测验了大学篮球队员，他们在空旷的场地（即没有任何防守者）上练习在15英寸处投篮。让这些球员针对100次投篮的命中率打赌。由于队员一般在这个距离上能够投中的比例约为50%，所以这样可以保证他们不一定能赢。打赌的规则是，当球员投中时赢的要比没投中时输的多一些。然而，球员可以选择下多些赌注（这样赢的多，输的也多）或者低的赔率（这样赢的少，输的也少）。显然的，如果球员能够预测自己的成绩的话，就会赢得比较多。这也就是说，当他们认为投中的概率高时，他们就会选择下多一些赌注；而当他们认为投中的概率低时，就会选择少下一点赌注。实验结果表明，就算是专业的球员也没有发现“有烫手”的现象：一次或多次投中后，再投中的概率并不比一次没中后再投中时更高。然而，球员们却都认为存在像“很烫手”这样的情况。他们在投中一球后，对下一次投篮所下的赌注，要高于在一次没投中后对下一次投篮所下的赌注。结果证明，球员们跟本不能预测自己的表现：他们预测的结果并不比随机预测的结果好。

赌徒的谬误来源于对概率这一概念的许多错误观念。其中的一个错觉就是，如果一个过程真正是随机的，就不可能出现有连续重复同样结果或模式的排列出现，哪怕是一个不起眼的随机事件（例如，掷6次硬币）。人们习惯地低估了在一个随机排列中，连续重复同样结果（HHHH）或样式（HHTTHHTTHHTT）的可能性。由于这个原因，人们在制作一组真正的随机排列时，却常常适得其反地制作出一个很少出现连续重复同样的结果或样式的排列。这是因为，人们在

制作时往往常常会错误地让可能的结果尽量轮流出现，以为这样才称得上是随机抽样，这样，当然就打乱了真正随机排列中所内含的这种重复结构（Lopes & Oden，1987；Nickerson，2002）。

那些声称自己有通灵能力的人可以轻易地利用人们这一错觉。让我们来看一看下面这个在大学心理学课上有时会进行的演示。让一名学生准备200个数字的排列。这200个数字是分别由1、2、3这3个数字中随机抽取的。完成之后，不要让作演示的老师看到。接下来，让这名学生集中精力在他的第一个数字上，老师则来猜这个数字是什么。当老师报告他的猜测之后，这个学生再向全班同学及那个老师公布正确的答案。以此类推，猜完这200个数字，也记录下老师猜对的数目。在实验开始之前，这个老师声称有通灵能力，可以在实验过程中可以用"通灵术"来得知别人的想法。通常，老师会先问班里的学生，他猜测的成绩会是怎样的——也就是"击中"的百分比会是多少——这样，他的演示才能成为支持他有通灵术的坚强实征证据。这时，通常都会有一个修过统计课程的学生会回答说，因为纯粹随机的猜测也都能得到猜中33%的结果，所以要想让别人相信他/她有通灵术，他/她猜中的比例就一定要超过33%，大约至少40%。班上大部分同学都会认同这一个观点。实验结束后，结果那位老师猜中的比例果真超过了40%。这个结果令很多同学感到惊讶。

学生们从这一演示中得到了一些关于什么是随机的教训，并且了解到伪装通灵能力是多么容易呀!在这个例子中，老师仅仅是利用了人们在制造那200个随机数字时，经常不让连续

重复的数字出现，以致于常在这3个数字中换来换去这一行为现象。在真正的随机排列中，已经出现了3个2之后，再出现2的概率是多少呢？其实还是1/3，与出现1或3的概率一样大。但大多数人在随机制作数字时并非如此。即使出现一个很小的两三个相同数字的重复，人们为了要得到一个“典型的”随机排列，常常会刻意地去交替给出不同的数字。这样，在我们的这个例子中，老师只要在每一轮猜测前，不去挑选那个学生在前一轮中挑选的那个数字，而从另外两个数字中选一个就可以了。例如，如果那个实验中的学生在上一轮说的数字是2，那么老师就会在下一轮的猜测中从1或3中任选一个。如果学生在上一轮说的数字是3，那么老师就会在下一轮的猜测中从1或2中任选一个。这样一个简单的把戏通常可以保证猜中的概率高于33%——高于3个数字随机猜测的准确率，而根本不需要动用什么通灵能力。

再谈统计与概率

以上所提到的这些问题，其实只是阻碍正确理解心理学的统计推理缺陷中的一小部分。有兴趣的读者可以参阅下面介绍的几本书，它们在这一方面提供了比较完整、详细的描述。这些书包括，由基诺维奇，格里芬（Griffin）和卡尼曼编写的《诀窍和偏见：直觉判断心理学，Heruistics and Biases: The Psychology of Intuitive Judgment》(2002)。另外，还有派艾尔蒂利-派尔玛瑞尼（Piattelli-Palmarini）的《不可避免的幻觉：推理错误如何统领我们的思维，Inevitable Illusions: How Mistakes of Reason Rule Our Minds》(1994)，拜瑞

（Baron）的《被误导的判断：大众抉择中的直觉和错误，Judgement Misguided : Intuition and Error in Public Decision Making》（1998），以及韦尔奇（Welch）的《抉择，Decisions》和《决策：有效抉择的艺术，Decisions: The Art of Effective Decision Making》（2002），吉格瑞恩的《精算过的冒险：如何察觉数字是在欺骗你，Calculated Risks: How to Know When Numbers Deceive You》（2002）。它们都对统计与概率作了很通俗的介绍（对没有受过任何数学训练的初学者尤其有用）。哈斯戴（Hastie）和达维（Dawe）的《在一个不确定世界的理性选择，Rational Choice in an Uncertain World》（2001）和拜瑞的《思考和决择，Thinking and Deciding》（2000）两本书中的介绍虽然有点专业化，但仍具有很高的可读性。

本章中所讨论的概率思维具有巨大的实践意义。由于没有充分运用概率思维能力，医生们选择了效果不好的治疗方法（Baron，1998；Dawes，2001）；人们不能准确地评估周围环境给他们带来的危险（Margolis，1996）；在诉讼程序中专家信息被误用（Foster & Huber，1999；Lees-Haley，1997）；政府和私人企业将数以百万计的资金用于不必要的项目（Arkes & Ayton，1999；Dawes，1988，pp.23～24）；动物不断被捕杀以至濒临灭绝（Baron，1998；Dawkins，1998）；不必要手术的实施（Dawes，1988，pp.73～75）；损失惨重的财政误判（Belsky & Gilovich，1999；Kahneman & Tversky，2000；Taler，1992；Zweig，2001）。

当然，我们在短短的一章里，不可能全面地讨论统计推

理。本章的目的仅仅是想强调统计在研究及理解心理学上的重要性。不幸的是，当碰到统计信息时，我们还找不到一个放诸四海皆准的规则。不像科学思维中的其他部分比较容易掌握，而统计学中，功能性的推理技术可能需要通过正规学习才能掌握。幸运的是，现在大多数综合大学和社区学院都提供了入门水平的统计学课程，这些统计学课程通常不需要以大学程度的数学课为其先修条件。在去上这类课程之前，读者可以先阅读我刚才推荐的那些书。

很多科学家都真诚地希望科学知识可以让一般大众也能知道及理解。然而，有时一门学科必须先掌握某些技术才能理解，而与这些技术相关的信息又只有在正规学习中才能获得，那么，如果我们让人们误以为对这门学科之深入理解，是一般外行人都能做得到的，那就是一个在学术上不负责任的做法。统计学和心理学就是属于这一类的学科。心理学家艾伦·班欧（Alan Boneau）调查了心理学教科书的作者，请他们列出学生在学习心理学时需要掌握的最重要的术语和概念（Boneau，1990）。在所列出的术语及概念中，频率最高的100个中，接近40%是属于统计学和方法论的范畴。现在已经到了不精通统计和概率的人，决不可能成为一个好的的心理学家的地步（Friedrich，Buday，& Kerr，2000）。

不可否认地，本书的一个目的就是要使心理学的研究能广为读者所接受。然而，在建构心理学的理论时所用的实证方法和技术，是和统计学相互缠绕到密不可分的地步（这一点，是和很多其他的领域一样，例如，经济学、社会学和遗传学等），以致于如果我们认为一个人可以没有一些统计学

的知识，就可以完全理解心理学，那就是大错特错了。因此，尽管这一章非常粗略地介绍了统计思维，但是它的主要目的是要强调要充分理解心理学，另外一门学科——统计学的知识是绝对不可少的。

小　　结

就像大多数的学科一样，心理学研究所得出的结论是概率式的结论——其所推论出来的结果是在大多数的情况下会发生的，也是较为普遍的，但这些结论绝不是在任何情况都会出现的。也就是说，它们不是100%准确的（就像所有医学的治疗方案那样，它们只是在一定概率的意义上有疗效）。尽管如此，根据心理学研究及理论所作出的预测仍然是有用的。

妨碍了很多人理解心理学研究的原因在于，他们很难用概率的术语来思考。在这一章里，讨论了几个相当精彩的研究例子，这些例子表明概率推理如何令大多数人止步不前：当人们遇到具体的、具有鲜活性的证据时，他们就将概率信息置之脑后了。他们也没有考虑到，较大的样本与较小的样本相比，它所估计出来的一些母体数值会比较更为精确一些。最后，人们也表现出赌徒的谬误：把原本无关的事件看成是有联系的。这个谬误是源于一个在下一章将要讨论的、人们更常有的一个惯性思维倾向：在决定因果关系时，未能认识到偶然机遇所起的作用。

（胡艳华　译）

第十一章　偶然性在心理学中的作用

上一章讨论了概率趋势、概率思维和统计推理的重要性。本章将继续这一讨论，但是把重点放在人们对随机性和偶然性这两个概念的理解困难上。我们将强调人们经常由于不能领会偶然性这一概念是如何彻底地融合在心理学理论之中，以致误解了研究对临床实践的贡献。

试图解释偶然事件的倾向

人类的大脑已经演化到一个程度，能让我们不懈地去寻求世界上的事物所存在的各种样式。对于身边发生的事物，我们总是想去寻找它们之间的关系、对它们进行解释并寻求这些事物背后的意义。这显然是一个极具生存适应的特点，它使我们能预测世界将会发生什么事，并且有可能最终可以找到为什么会出现某一样式的精确解释。当然，这种对样式的探求是科学家的工作，但它并不局限于科学家，我们每个人在生活中的每一天，都在进行着这种探寻活动。心理学家曾经研究了人们这种探寻事物结构的强烈倾向，并指出它是人类智力的特性，而且这一活动是人类在信息加工和知识获得方面，能取得许多令人震惊的伟大成就背后之真正原因。

然而，人类认识过程的这种极具生存适应性的特点并非十全十美，有时还会给我们带来一些障碍。例如，当环境中没有什么东西可以让人们进行概念化时，还一味地去寻求用概念来理解事物，就是一种不良的适应行为。那么，到底是什么东西在人类认知的这一突出特点上制造麻烦呢？又是什么东西给我们寻求结构这一过程带来困扰，以致于阻碍了我们对事物的理解呢？对，你猜对啦！是概率。更具体的说：是偶然性和随机性。

偶然性和随机性是我们周围环境不可分割的一部分。偶然性和随机性的规律是支配生物进化和基因重组的机制，物理学也运用关于偶然性的统计定律来解释物质的基本结构。自然界发生的很多事情都是稳定的、系统的、可解释的因素与偶然因素综合作用的结果。回想前面谈到的一个例子：吸烟导致肺癌。生物学上系统的、可解释的因素把吸烟和某一疾病之间建立联系，但这并不表示所有吸烟者都会患肺癌，这种趋势是概率性的。也许，最终我们能解释为什么有些吸烟者不会患肺癌，但是在现阶段，这种不稳定性必须归因于决定一个人是否患某一疾病的大量偶然因素。

上例中说明一件事决定于偶然性时，并不必然表示它是不确定的，而只是说它是当前无法确定的（Alcock，1981，p.155）。例如，说抛掷一枚硬币是偶然事件，并不是因为当通过测量抛掷的角度、硬币内的金属含量以及许多其他变量等，都不可能确定其抛掷的结果，实际上，这些变量的确决定了抛掷硬币的结果。但是，正是因为在每一次抛掷时，我们没有比较简易快捷的方法来测量这些变量，所以才将之视为是一个偶然事件，一次抛掷的结果不是具有不确定性，它只是在当下是无法确定

的而已。

世界上的许多事件不能只运用稳定的、系统的因素来完全解释清楚，至少现在还不能，我们必须求助于在事件中也起作用的许多偶然因素。不幸的是，大多数人都觉得这样做不能令人满意。当然如果这种质疑激励了我们对一种现象去做进一步的探究，这是一件好事。但是如果它促使我们匆匆忙忙地去接受某一种解释，特别是当这种解释尚缺乏实征验证的支持，就已经声称它完全可以解释该事件时，就会给我们造成阻碍。

尽管，有时我们暂时不能找到对一个特定现象的系统解释，但是我们的不断寻求概念化的认知加工系统仍然在不停地工作，企图把无意义的理论强加于原本随机的数据。心理学家曾对此现象进行了实验研究。研究者在一个实验情境中，要求被试观察在许多不同维度上变化的一系列刺激物，并告诉他们其中的一些刺激物是属于一类，而其他的则属于另一类，被试的任务是去猜测每一个刺激物是属于这两类中的哪一类。实际上，研究者是把刺激物随机归类，因此，除了随机性，并没有任何其他规律。但是，被试很少敢做随机猜测，相反地，他们通常处心积虑的去虚构一个复杂的分类方法来解释他们给这些刺激物所做的分类。

同样的，各种各样的“预谋论（conspiracy theories）”通常也需要搞出一层又一层的复杂计谋，来帮助那些积极寻求一个稳定解释的预谋论者，把一个原本是随机出现的一连串孤立事件，说成是有预谋的相关事件。例如，许多金融分析者的思维方式就体现了这一谬误。他们通常对股票市场价格每一次小小的波动都作出精心的解释，而实际上这种波动大多只是随机波

动而已（Malkiel，1999；Taleb，2001）。然而，股票市场分析者总是让人感觉他们可以（也许他们也相信自己可以）“征服市场”。其实，当大量的证据表明，他们是做不到的。如果你购买了包括在标准普尔（the Standard and Poor's Index）500 指数中的大企业之股票，然后存放起来，不去理它（可称之为“傻子策略”），经过了 20 世纪 70 年代的整整 10 年，其可获得的回报率，会比华尔街 80% 的股票经纪人为他们的顾客所赚的回报率还高（Malkiel，1999）；如果封存是在 80 年代的 10 年里，你可以打败华尔街 2/3 的经纪人（Malkiel，1999；Updegrave，1995），也可以打败 80% 的那些每年要订阅者出 500 美元才能看得见的财经通讯刊物（Kim，1994）。

但是，话又说回来了，我们要如何来看待那些确实打败了傻子策略的经纪人呢？你可能会问，这是不是表示他们具有某些特殊的才能。要回答这一问题，让我们先设想下面这一个实验：有 100 只猴子，每只猴子手中握有 10 支飞镖，它们都向一面写有标准普尔500指数中的大企业名字的墙上投射这些飞镖，飞镖投中的地方上所写的企业就代表猴子那年要买的股票。那么，一年后它们的业绩会是怎么样的呢？有多少只猴子会打败全买标准普尔 500 指数股票的业绩？对了，你答对了!大概有一半的猴子会。那么，你会不会愿意付钱给这一半打败标准普尔 500 指数的猴子，要它们下一年帮你炒股呢？

有关财经预测的例子还可以帮助我们进一步说明，纯粹随机发生的一连串事件，是通过什么逻辑会被看起来像是由可以预测的因素所造成的。接下来要举的这一例子引自保罗斯（Paulos，1988）和佛雷德森（Fridson，1993）。假想你收到一

封信告诉你有一份关于股票市场预测的邮寄通讯已经问世，这个通讯并不是要你订阅，只是要求你试试照着他们的建议去买股票，然后看看它的预测灵不灵。它告诉你IBM的股票会在下个月攀升。你把这一通讯放在了一边，但是你确实注意到在下面的一个月里IBM股票果真涨了上去。如果你曾读过一本与本书的内容相似的书，你可能就知道，这并不代表什么，没有什么了不起。你仅会用粉笔记下这一侥幸的猜测。后来你又收到另一份来自同一家投资咨询公司的通讯，该通讯说IBM会在下月下跌，当股票确实下跌了时，你又一次用粉笔记下这一侥幸的猜测，但是，这一次你可能就有些好奇了。当这家公司寄来第3份通讯，预测IBM下个月会再次下跌时，你发现自己会更密切地关注财经版的新闻，你发现这个通讯又一次地作了准确预测，IBM这个月确实又下跌了。当来自这家公司的第4份通讯说IBM下月会攀升，而也确实攀升了时，你就很难不会去想，这个通讯刊物还真有点本事——很难摆脱想要花29.95美元去订阅一年的冲动；惟恐这个通讯会在第四次免费赠阅之后就不再寄来了。是的，如果你不知道这里面的秘密的话，确实很难摆脱这种冲动。但是，如果你能想像，有一个人正在一个简陋的地下室办公室内，将下周要发出去的1600份通讯，其中800份预测IBM下月攀升，800份预测下跌，按电话簿上找到的1600个地址，寄了出去。那么，一切就不足以为奇了，当IBM在下个月真的攀升了，公司就继续把通讯只发给上月接收到正确预测的800位“客户”（当然，其中还是400份预测攀升，另外400份预测下跌）。然后，你可以继续想像，这个“锅炉房”——可能与电话市场的炒作骗子共谋，

利用电话来进一步跟进这一骗局——正在向第2周接收到正确预测的400客户发送第3个月的预测通讯。接下来，对了，你就是连续四次收到正确的随机预测信息的一百个幸运者之一（多么令人激奋呀！）同时你也是那一百个开始相信该通讯的准确度，从而会被进一步要求，付29.95美元来订阅它的幸运儿之一。

知道了内情之后，我们看到这是一个多么可怕的骗局啊！确实是这样的。但是，这比“受人尊敬”的财经杂志里或电视节目上所展示给你看的“股票经纪人连续四年打败多半对手”的情节要好得多。请回想一下猴子投镖的情境，设想这些猴子是财务经纪人，年复一年的帮人选股。很明显地，第一年，他们之中有50%会打败他们的对手。第二年，这50%的人中又有一半——按随机概率来算，打败他们的对手，即只剩下25%的经纪人能连续两年打败他们的对手。之后第三年又有一半——按随机概率来算，打败对手，即只剩下总人数的12.5%连续三年打败对手。最终到第四年，又只有这些人的一半（总人数的6.25%）能打败自己的对手。因此，100只猴子中只有6只能取得像财经节目和报纸所说的“在连续四年里打败了其他的经纪人”的骄人成绩。那么，既然只有这6只猴子能打败了所有其他投镖猴子（正如前面所看到的，也就是打败了大多数“现实生活中的”华尔街经纪人，见Malkiel，1999），它们当然可以理直气壮地在“华尔街一周”电视节目中亮相啰，你不觉得吗？

解释偶然所造成的虚幻联系和控制错觉

人们喜欢对偶然事件进行解释的行为，在心理学家研究的一个称为虚幻联系现象（illusory correlation）中看得最清楚。当人们相信两类事件通常应该一起发生时，他们就会认为，这两个事件一起出现的频率是比较高的，即使这两类事件的同时出现是随机发生的，而且并不比任何其他两个事件同时发生的次数为多。总之，人们甚至在随机事件中，也倾向于见到他们所预期的联系（Nisbett & Ross，1980；Stanovich，1999，2004）。他们在本来没有任何结构的一些事件中看到了虚幻的结构。

有控制的实验研究结果显示，当人们已经预设有两变量相联系的想法时，即使在两个变量根本毫无关系的数据中，他们也可以找到联系（如 Broniarczyk & Alba，1994；King & Koehler，2000；Stanovich & West，1998）。不幸的是，这一现象在现实世界中也广泛地出现，以至于给人们的生活带来了许多负面的影响。例如，许多从事心理治疗工作的人都一直相信罗夏克（Rorschach）测验是有效的。这个著名的墨迹测验要求被试对一张白纸上的墨迹做出反映。因为这一墨迹没有结构，所以这一测验的发明者认为，人们会以自己对模糊情境的定型反应来对这些墨迹做出反应，从而能够揭露其“潜藏的”心理特质。这种测验也被称为是投射式（projective）测验，因为发明者认为被试对墨迹做出的反应，应该是他们下意识的心理、思维和感受的投射。然而，问题是，没有任何证据表明罗夏克测验作为一个投射测验，提供了任何额外的诊断价值

（Dawes，1994；Garb，Florio，& Grove，1998；Lilienfeld，1999；Lilienfeld 等，2000；Shontz & Green，1992；Wood，Nezworski，& Stejskal，1996）。对罗夏克测验的信心可以说是起源于虚幻联系这一现象。临床心理医生从病人的反应定型中找到了联系，因为他们早已相信这些联系是存在的，而不是因为真的观察到了什么联系。

心理学家雷·海曼（Ray Hyman）讨论了人们喜欢在原本没有联系的事件中寻找联系的倾向：

为了要理解世界上的事物，我们必须运用自己原有的知识和预期。在大多数的一般情境中，利用我们的知识背景和记忆，可以令我们正确地去解释对世界的观察，并为这些解释提供必要的推论。但这一有力的机制，在本来根本没有任何信息需要传递的地方就会走歪了，我们会错误地、不断地去追求原本没有意义的意义，而不去把由随机性所造成的噪音给丢掉（Hyman，1981，p.96）。

在我们的生活中，许多人与人之间的交往都有很大的偶然成分在里面：“双方不认识的男女约会最终促成了婚姻；取消了的约谈导致了工作的丧失；由于误了班车而遇到了高中的老同学等等。认为生活中每一件偶然的小事都需要认真地去解释的想法是错误的。但是，如果偶然事件确实产生了重要的后果时，人们不免要建构一些复杂的理论去解释它们。”

试图去解释偶然事件的倾向可能源于我们有深切的愿望想要相信自己是可以控制这些事件的。心理学家埃伦·兰格（Ellen Langer）研究了控制错觉（illusion of control）这一现

象。它是指，相信个人的能力能够影响偶然事件的结果。在一个研究中，两个不同公司的雇员卖彩票给他们的同事，对其中一些同事，他们只是简单的派发他们每人一张彩票，而对另一些同事则让他们自行抽取一张。当然，在随机抽奖的事件中，无论彩票是在给予他们选择权利的情况下买的，还是被硬性派发的，都没有区别，中奖率都是一样的。但是，第二天，当这两个雇员试图向这些同事买回那些彩票时，给予选择权利的被试对彩票的要价是硬性派发的被试要价的4倍。在另外几个实验研究中，兰格证实了造成这一结果的原因在于，人们不能接受个人的能力无法对偶然事件产生影响这一事实。这一控制错觉在美国所有各州发行的彩票买卖中都有出现，表明了它是相当普遍的现象。这一现象自然为那些教导人们如何“战胜”彩票的伪科学书籍开了方便之门，它们充斥市场（Clotfelter & Cook，1989）。这类书之所以畅销，是因为人们不懂得什么叫随机，其意义又是什么。事实上，美国是到了20世纪70年代中期各州才爆发出购买彩票的热潮，导火线就在于新泽西州发明了一种新的售卖彩票方式，那就是，让购买者可以自行刮票中奖，或由他们自行在彩票上挑选自己所要的数字（Clotfeler & Cook，1989；Thaler，1992，p.138）。用这种售卖方式来进行的抽奖活动通常叫作参与性抽彩（participatory games），而这类参与性抽彩正是利用了当时兰格刚研究出来的控制错觉现象：人们错误地相信他们的参与行为能够决定随机事件，让他们有多点儿机会中奖。

另外有一些心理学家则研究了另一个与此相关的现象，被称为是公平世界假设（just-world hypothesis），它是指人们倾向

于相信自己是生活在一个公平的世界里，每个人都得到了他/她所应得的报应（Lerner & Miller，1978）。研究者证实了，包括在这一公平世界假设中的一个信念，会导致人们对因偶然所造成的不幸事件之受害者不予同情，认为他们是罪有应得。倾向于为偶然事件寻求解释正是产生不同情这类受害者现象的原因，人们很难相信一个完美无邪的或是道德修养好的人会因为偶然事件而惨遭不幸。固然我们想要相信好人有好报、坏人有坏报，但是，偶然却是完全不带偏见的，所有人都有同等的概率得到好的或是坏的偶然事件。

公平世界假设这一信念当被推到极端时，会产生出一些有害的或是不人道的教条。请看，在80年代早期，美国教育部的一个官员所陈述的观点背后之逻辑，就可略见一斑了。他说，残疾人“错误地认为生命中的巧合，令他们随机地受到惩罚，其实这是错误的。没有什么发生在一个人身上的事不是他本人，在他发展过程的某一点上，所招致的……这句话听起来好像很不公平，但是一个人所处的外部环境确实是与他内在心灵的发展是相一致的”（Gilovich，1991，p.143）。正如基诺维奇所指出的，“这真不该是一个想进入教育部高层的官员所持有的哲学，教育部本应是负责给予残疾人士同等教育机会的机构才对”（pp.143～144）——但是这正是拒绝承认偶然这一因素在起作用之不可避免的结果。

体现在公平世界假设中，这一对偶然的错误理解，也可以用来理解许多其他不正确的民间信念，这些信念导致我们见到虚幻联系的作用。例如，在第六章提到，盲人有非常敏锐的听觉，这就是一个不正确的信念。然而，这一个民间信念之所以

永远流传，可能是因为人们想要看到这个联系，这样可以令他们觉得盲人得到了应有的补偿(见Niemeyer & Starlinger, 1981; Stankov & Spilsbury，1978)。

偶然和心理学

在心理学中，研究者也有试图去解释任何事件的习惯。他们不仅只对行为中系统的、非随机的事件要提出理论进行解释，而且也不放过其他变化，以及任何细小的差异总会有人提出一个理论来解释它们。这一现象的结果是造成不可证伪理论的泛滥，这些理论包括那些仅只是个人提出的、未经公开检验的理论和那些伪科学的理论。“心理历史”研究法的奉行者，经常犯了这一错误，一个著名人物生命中的每一个细小的变化及转折，通常通过精神分析学派的理论，都被解释成为是重大的心理历史事件。而这些心理历史事件的问题，不在于它们解释得太少，而是它们解释得太多。这一研究方法的奉行者很少承认决定一个人的一生的因素有许多是偶然的。

对于要想使用心理学知识的外行人来说，理解偶然这一因素的作用是非常重要的。受过正统训练的心理学家承认他们的理论只能解释人类行为变化的一部分，而不是全部，他们会坦率地接受偶然因素的作用。但是，那个在温佛瑞秀电视节目中出现的嘉宾（见第四章开头）则几乎针对每一个个案，以及人类行为的每一个小节都能用他的理论提供一个解释。我们对他的这一天分，应该产生怀疑，而不是尊敬。真正的科学家从不惧怕承认他们的无知。总之，评价心理学说

法的另一实用法宝就是：在接受对一个事件的复杂解释以前，先想一想，这一事件的哪一部分事实上只是由偶然因素所引致的。

巧合

对实质上是偶然事件进行解释的习惯，也导致我们对许多巧合事件（coincidence）的性质产生误解。许多人认为巧合需要特别的解释，他们不理解巧合这一现象，即便只是由偶然所引起的，也是必定会发生的，并不需要特别的解释。

《韦氏新世界字典》（*Webster's New World Dictionary*）里把巧合定义为：“相互关联或相同的事件意外地、令人不可思议地同时出现了。”因为在同一本字典里，把意外的（accidental）定义为“因为偶然（chance）而发生的”，所以毫无问题的是，巧合只是相关事物基于偶然的因素才同时发生的。不幸的是，许多人并不这样解释巧合，自动去寻找事件之间的样式和意义的习惯与巧合的“不可思议的”这一性质结合在一起，使得许多人忘记了他们可以用偶然这一因素来解释巧合。相反地，他们却精心地去炮制一些理论来解释巧合。下面讲的这个故事你一定已经听过不少遍了：“那天我坐在那里，正在想好久没有叫比尔叔叔到得克萨斯州来玩玩了，紧接着电话铃就响了，你猜怎么着？是的，没错！正是我老比尔叔叔打来的，看来还真有“‘心电感应’这玩意儿!”这是精心地去解释巧合事件的一个典型例子。每天，我们大多数人都可能想到很多不同的、住在远方的、认得的人，这些人在我们想起他们时，有多少人可能会打电话来呢？几乎没有可能。因此，在一

年内，我们可能想到了数百个人，而这些人中没有一个曾打电话来，但是我们并不介意，因为这是常态。最终，在这数百个我们从未意识到的“否定事件”之后，一个罕见的事件发生了——当我们想到某一个人时，他/她正好打电话来了！我们可以看到，这件事是在数百件同样的事件中很难得出现的一个情况。然而难得的事件是会出现的，而且其出现纯粹是偶然，其他解释都是不必要的。

如果人们确实理解巧合的意思（一个因偶然而发生的令人不可思议的事件），他们就不会掉入要去寻找系统的、非偶然性之解释的陷阱。但事实正相反，巧合对很多人来说，都不被认为是偶然的结果，而是需要解释的。例如，许多人都听到过这样的一句话“天哪！这真是太巧了！怎么会这么巧呢？”为此，马科斯和卡曼（Marks & Kammann，1980）建议大家今后用一个比较中性的名词像奇配（oddmatch）来表示某两件事令我们感到非常奇怪地、以及不可思议地同时发生的现象了。奇配这一名词并未具体指明两件事同时发生是基于偶然，还是基于系统的解释。

认为罕见的事不会发生，认为奇配绝不是因为偶然，这类错误观念是我们之所以要为巧合事件寻求解释的一个原因。我们的这类错误信念之所以这么强烈，也是因为概率有时是用对比率（odds）这一概念来表述的，而这种表述的言外之意起到了一定的作用。看看我们是用什么方法来叙述概率的：“啊！天哪，这事件是很不可能发生的！因为它出现与不出现的对比率是1∶100!”我们在做这样的表述时所用的方式强烈地让人感觉这件事绝不会发生。当然，我们可以用另外一种

表达方式来表述同一件事，而这一方式可能给人带来完全不同的感受：“在100个同类的事件中，这样的结果可能会发生一次。”这种表述方式强调，尽管这一事件是少见的，但是长时间来看，罕见的事终究一定会发生的。简言之，奇配是会因纯偶然而发生的。

实际上，概率定律保证，随着事件发生次数的增加，一些奇配出现的可能性变得很大。也就是说，偶然事件的规律不仅让奇配迟早会出现，而且长时间来看，它几乎可以说是保证了它一定会出现。请看马科斯和卡曼（1980）的例子，如果一次掷5枚硬币，结果它们都是正面朝上，你将认为这是一个奇配、一件不可能发生的事。是的，它发生的概率是1/32或.03。但是如果你掷100次，再问：在这100次中，至少有一次全部正面朝上的可能性是多少呢？答案是.96，就是说，100次中，这个罕见的奇配是非常可能发生的。

因为美国很多州都可以卖彩票，通过电脑或其他装置随机抽出中奖号码，所以每周的摇奖活动累积下来，不可避免地会有很不可能的巧合数字出现，例如连续两周中奖号码是相同的，这一奇配的出现虽然只会引起许多统计学家和行为科学家的一两声哼笑，觉得是一个奇妙的巧合，但是这一巧合却会引起公众的抗议，认为这表明摇奖过程有作弊或动了手脚的嫌疑（见Clotfelter & Cook，1993）。公众之所以会认为彩票公司有人作弊，起因于他们错误地认为，像这样罕见的或不大可能出现的事情，不可能会仅仅是偶然因素的结果。当然，统计学家哼笑的理由是，偶然恰恰是以与公众的逻辑相反的方向在起作用，如果彩票发行了一段长时间，连续出现相同中奖号码的事必然

会发生！

若干年前，安 · 兰德（Ann Lander）* 宣传了一系列有关亚伯拉罕·林肯（Abraham Lincoln）和约翰·肯尼迪（John Kennedy）总统之间令人“毛骨悚然”的巧合：

1. 林肯于1860年当选总统；肯尼迪则于1960年当选。
2. 林肯和肯尼迪都关注民权。
3. 林肯和肯尼迪这两个名字都有7个字母。
4. 林肯有一个叫肯尼迪的秘书，肯尼迪也有一个叫林肯的秘书。
5. 两人都由叫约翰逊的南方人继任。
6. 两人都被用三个名字的人暗杀（John Wilkes Booth和Lee Harvey Oswald）。
7. Booth 和 Oswald 都持有不受人欢迎的政见。
8. Booth 在电影院里射杀了林肯然后藏在仓库里；Oswald 从仓库中射杀了肯尼迪然后藏在电影院里。

当然，作为巧合，这些事之间的联系并不是令人毛骨悚然的。得克萨斯大学的电脑程序师约翰 · 李维（John Leavy）曾经搞了一个叫“令人背脊发凉的总统巧合竞赛”来说明，实际上在任何两个总统之间找到像上述那样的一个列单是多么容易的事（见 Dudley，1998）。例如，李维（Leavy，1992）的文章中包括了对威廉 · 亨利 · 哈里森（William Henry Harrison）和扎卡里 · 泰勒（Zachary Taylor）的比较，波尔克（Polk）和卡

* 安·兰德为美国已故的报纸专栏作家，以回答读者有关生活礼节及事故的问题见长。

特（Carter），加菲尔德（Garfield）和麦金利（McKinley），林肯（Lincoln）和杰克逊（Jackson），尼克松（Nixon）和杰斐逊（Jefferson），华盛顿（Washington）和艾森豪威尔（Eisenhower），格兰特（Grant）和尼克松（Nixon），麦迪逊（Madison）和威尔逊（Wilson）等之间的相似之处。下面是关于加菲尔德（Garfield）和麦金利（McKinley）之间惊人的相似处：

1. 他们俩都在俄亥俄州出生长大。
2. 他们俩都是美国内战的老兵。
3. 他们俩都在众议院任过职。
4. 他们俩为了保护美国工业，都支持金本位制和关税保护制。
5. 他们俩的名字都有8个字母。
6. 他们俩都被来自纽约市的副总统取代（西奥多·罗斯福，*Theodore Roosevelt*和切斯特·亚伦·阿瑟，*Chester Alan Arthur*）。
7. 罗斯福和阿瑟两个名字都有17个字母。
8. 两个副总统都蓄须。
9. 他们俩都在任期第一年的9月被枪杀。
10. 他们的刺客，查尔斯·基埃图（Charles Guiteau）和利昂·周格资（Leon Czolgosz）听起来都不像美国人的名字。

许多关于总统之间联系的列单都很类似。总之，在一个人的一生的数十年中，由于人际交往及各类事件的复杂性，在内藏着成千上万事件的取样空间里，任何两个人之间，如

果没有找到什么相似之处，那才是让人奇怪哪（Martin，1998）。

懂得在什么时候，抑制对随机事件编造复杂的解释，是具有实际意义的。作家艾图尔·嘉万迪（Atul Gawande）曾描述了1973年赎罪日（Yom Kippur）战争期间，认知心理学家卡尼曼因提建议而遭到以色列空军部队批评的故事。两个飞机中队出发了，回来后，一队损失了4架飞机，另一队没有损失。部队希望卡尼曼调查一下，之所以有这样的差异结果是否是因为有特别的因素在起作用。卡尼曼并没有去做调查，他仅仅运用了本章所谈到的逻辑去告诉以色列空军部队不要浪费时间："卡尼曼知道如果空军长官真的去调查，他们将不可避免地发现一些可测量到的差异，以致使自己非要做一些举措来弥补这一差异"（Gawande，1999，p.37）。但是卡尼曼知道，任何找到的因素都极有可能是虚幻的——只是偶然波动的结果而已。

个人的巧合

在我们个人的生活中，奇配的发生往往因为它们对我们具有特殊的意义，而更不易归因于偶然了。产生这种倾向的原因有很多，某些是动机性和情感性的，但还有一些是概率推理的失败。我们通常不能意识到，奇配只是包括了众多"平凡"事件的一个大事件库的一个非常小的部分而已。对某些人来说，奇配看起来好像经常发生，但是它真的是经常发生吗？

请想一下，如果我们现在去分析在你个人生活中的奇配，其结果会告诉我们什么呢？下面的例子引自马科斯和卡曼

(1980)。假定在某一天里，你参与的不同事件有100个，这个数目其实并不算多,设想一下我们是生活在一个多么复杂的现代工业社会！实际上，100个事件可能还少了一点儿。你看电视、打电话、与人面谈、商定去工作或去商场的路线、做烦人的家务、阅读吸收信息、在上班时完成复杂的任务等等。所有这些事件都包含很多成分、而各成分又可以被看成是一个个可以分开来记下的小事件，这样加起来，说一天中一个人涉及100个事件，其实真不算多。不过，我们就拿100个事件来作例子吧。奇配是指其中两个事件不可思议地联系在一起了。那么典型的一天中这100件事之间共有多少不同的、两两匹配的组合呢？用一个简单的公式来算一下，我们就可以得到，一天中有4950个不同的两两匹配，一年有365天，天天如此。我们知道，奇配是令人难忘的，比尔叔叔打来电话的那一天，你可能几年不忘。假定你把在10年期间所记得的所有奇配数出来，也许也就六七件(或多或少，不同的人对奇配一词有不同的标准)。这六七件是从一个多大的平凡事件库中出现的呢？每天4950个两两匹配，乘以一年365天，再乘以10年，得18067500个匹配。总之，10年中如果有6个你认为是奇配的联系发生了，就有18067494个其他两两匹配，但也可能是奇配的联系也发生了。所以，在你的生活中的一个奇配，其发生的概率是.00000033！这样，在1800万个事件联结中，只有6个是奇配，其实应该是不足为奇的了。罕见的事件确实发生了，它们也的确是少见的，但是偶然这一因素保证了它们一定会发生(请回想前面掷5枚硬币的例子)。在我们的例子中，在你的身上发生了6件奇事，它们可能是巧合：两个相关事件基于偶然不可

思议的同时发生。

心理学家、统计学家以及其他科学家都指出，人们认为许多奇配的少见程度是比实际上它们出现的频率要少得多。在一篇副标题为“为什么巧合的长臂没有看起来那么长？”的文章中，作者马丁·加德纳讨论了许多奇配并不是像通常所想像的那样罕见这一事实（Gardner，1972）。著名的“生日问题”是最好的例子。在一个23人的班级里，有两个人同一天过生日的概率是多少？大多数人会认为非常低。而实际上，23人的班级中，2人同一天过生日的可能性是50%！而35人的班级，可能性就更大了（概率大于.80；见Martin，1998）。所以，因为美国历史上有41位总统，其中詹姆斯·波尔克（James Polk）和沃伦·哈丁（Warren Harding）两位是在同一天出生（11月2日）这一事件联系，就没有什么了不起了。同样的，有36位总统都已过世，其中米勒德·菲尔莫尔（Millard Fillmore）和威廉·H·塔夫脱（William Howard Taft）两位死于同一天（3月8日），也不应令人惊讶，甚至还有另外3位总统，约翰·亚当斯（John Adams）、托马斯·杰菲逊（Thomas Jefferson）、詹姆斯·门罗（James Monroe）都死于同一天，而这一天竟然是7月4日，美国国庆节！你说奇不奇怪？其实不奇怪——概率使然而已。

这里所要表达的意思是，很多奇配其实并不奇特。心理学家鲁马·范克（Ruma Falk）和唐·麦克格瑞格（Don MacGregor）研究了人们如何判断一件事是否值得感到惊讶（Falk & MacGregor，1983；Falk，1989）。他们研究的一个重要发现是，在判断巧合事件时，我们是很自我中心的：人

们觉得发生在自己身上的巧合比发生在别人身上同样的巧合，要来得更令他们惊讶。范克和麦克格瑞格（1983）推测这一结果的原因可能是“我们常常认为自己是独特的、跟别人不一样的，因而看不到发生在自己身上的事情，同样地也可能发生在别人身上。换句话说，我们可能不容易看到，在一个事件或情境中的自己，是可以换成为另外一个人的”（p.501）。

接受错误以减少错误：临床式与精算式预测

人们在试图解释世界上发生的事时，往往不愿意承认偶然因素的作用，这一现象实际上降低了我们对现实世界的预测能力。在某个领域中，承认偶然因素的作用意味着研究者必须接受一个事实，那就是，我们的预测不可能百分之百准确，我们总是会犯错误。但有趣的是，承认我们的预测达不到百分之百的准确度，实际上反而提高了整体预测的准确性。这听起来好像有点儿矛盾，但是事实确是如此：为了减少错误就必须接受错误（Dawes，1991；Einhorn，1986）。

“接受错误以减少错误”这一概念可以通过在认知心理学实验室里已经研究了数十年的一个非常简单的实验任务来说明之（Fantino & Esfandiari，2002；Gal & Baron，1996）。这个实验任务是这样子的，被试被安排坐在两盏灯（一红一蓝）前，实验者要求他们去预测每次实验测试时哪一盏灯会亮，被试要参与很多次这样的测试，并按准确率给予一定的报酬。实际上，所有的测试都是在70%的次数亮红灯、30%的次数亮蓝灯

的条件下进行的，两种灯是以随机顺序出现。实验过程中，被试很快就感到红灯亮的次数比较多，因此也就在更多的测试中预测红灯会亮。事实上，他们确实有在大约70%的测试中预测红灯会亮。然而，正如前面所讨论的，被试在实验过程中逐渐发现并相信灯亮是有一定样式的，但却从没想过每一次哪一个灯会亮本身却是随机的。因此，为了要使他们的预测百发百中，他们就在红灯与蓝灯之间换选，但保持70%的次数预测红灯会亮，30%预测蓝灯会亮。被试极少意识到，如果他们放弃力求自己”百发百中”，他们的预测会更好一些！为什么会是这样的呢？

让我们思考一下这一情境背后的逻辑。在以70：30的比例随机点亮红灯或蓝灯的条件下，如果被试在70%的测试中预测红灯会亮，30%的测试中预测蓝灯会亮，他的准确率会是多少呢？我们将用实验中间部分的100个测试来计算——因为那时被试已经注意到红灯亮的次数比蓝灯多，从而开始在70%的测试中预测红灯会亮了。因为在100次测试中有70次红灯亮了，被试在这70次中有70%的准确率（因为被试在70%的测试中预测红灯会亮），也就是说，被试在70次中有49次正确的预测；100次测试中有30次蓝灯亮了，被试在这30次中有30%的准确率（因为被试在30%的测试中预测蓝灯会亮），也就是说，被试在30次中有9次正确的预测。因而，在100次测试中，被试的正确预测是58次（49次加9次）。但是，请注意，这是多么可怜的成绩啊！如果在被试注意到哪一盏灯亮得比较多后，就总是预测是那盏灯会亮——在本实验中，就是注意到红灯亮的次数比较多，据此就总是预测红灯会亮（姑

且把它称之为“百分之百红灯策略”），那么，他在100次测试中会有70次正确的预测。虽然在蓝灯亮的30次测试里，被试将没有一次正确的预测，但是总准确率仍然高达70%——比在红灯与蓝灯之间换选，以追求“百发百中”之58%的准确率要高12个百分点。

然而，百分之百红灯策略所取得的、较高的准确率是要付出代价的：必须放弃要“百发百中”的雄心。（显然的，虽然明知有时蓝灯会亮，被试还总是预测红灯会亮，本身已表明他已经失去在蓝灯亮的测试中命中的机会）。这就是接受错误以减少错误！放弃在预测每一单一事件时不犯错误的想法，会让被试获得更高的总体准确度。同理，如果在预测人类的行为时，有一定的准确度，有时就要接受错误以减少错误，也就是，在依靠一般性的原则来做出比较准确的预测的同时，也要承认我们是不能在每一个单一事件上都预测准确的。

但是，接受错误以减少错误，说来容易，做起来难。心理学这一领域里，人们花了40年时间，来研究临床式预测和精算式预测这一课题，就是有关“做起来难”的最好例子。精算式预测（actuarial prediction）是指依据由统计资料中分析出来的群体趋势，所作的预测。在本章一开始时所讨论的群体（也就是总体）预测就是属于这种预测。一种简单的精算式预测是，针对凡是具有某种特征的所有个体，做出相同的结果预测。例如，预测不吸烟者的寿命是77.5岁，而吸烟的人是64.3岁就是一个精算式的预测。如果考虑的群体特征不只一个（运用第五章谈到的复杂相关技术——尤其是多元回归技术）将令我们的预测更加准确。例如，预测吸烟、肥胖且不作运动者的寿命是58.2

岁就是在一个多变量（吸烟行为、体重和运动数量）基础上的精算式预测，这样的预测总是比单变量的预测更准确。精算式预测在经济学、人力资源、犯罪学、商业与市场学，以及医学科学等领域都很常见。

在心理学的许多分支领域，如认知心理学、发展心理学、组织心理学、人格心理学与社会心理学中，其知识都是通过精算式预测来表达的。相反的，有一小群在外面执业的临床心理学家，则声称他们可以超越群体预测，能对个别个体各自做出百分之百准确的预测，这种预测被称为是临床式或个案式预测（clinical or case prediction）。与精算式预测相对比，临床式预测是这样的：

职业心理学家声称，他们能对具体个体进行预测，从而超越了对“一般人”或不同类别的人所进行的预测……职业心理学家与心理学研究者的不同在于，他们能把每个人看作为独特的个体来理解，而不仅只是一个群体的一员，之所以只能作群体理解是因为统计学的概化必须是针对一群人来作的。职业心理学家自认为能够分析出，在一个个体的生活中，到底是“什么导致了什么”，而不只是说说什么样的人“总体而言”是这个样子的。（Dawes，1994，pp.79～80）

临床式预测应该可以是精算式预测的一个有用的补充，但问题是，临床式预测是完全不准确的。

要证明临床式预测是有效的，一个临床医生与他的病人接触的经验，以及他/她对病人所提供的信息之运用，应该会给我们一个比较好的预测。这个预测一定要比只是拿对这些病人

的一般资料事先进行的简单量化编码,然后放在一定的统计程序中去找出这些资料的最佳预测组合的研究结果，来得好才行。总之，“临床式预测是有效的”这一观点认为，临床心理学工作者的经验会使他们超越精算式预测研究所能提供的总体联系信息。既然讲得这么明了，那么，就很容易可以检验一下这一观点的正确性。很不幸的是，经过检验，这一观点被证明是错误的。

对临床式与精算式预测的比较研究所得到的结果，具有惊人的一致性。自从保罗·米尔（Paul Meehl）的经典著作《临床式预测对精算式预测》于1954年出版以来，40年间有100多个研究表明,在几乎每一个曾经进行过研究的临床式预测领域（心理疗法的效果、假释行为、大学生毕业比例、电击治疗的疗效、累犯问题、精神病住院治疗期的长短等等），精算式预测都优于临床式预测（Dawes，1994；Dawes，Faust & Meehl，1989；Faust，Hart，Guilmette，& Arkes，1988；Goldberg，1959，1968，1991；Leli & Filskov，1981；Ruscio，2002）。

在多个临床领域中，研究者给予一个临床心理医生有关一个病人的资料，让其预测这个病人的某些行为。与此同时，他们也把同样的资料加以量化，并用一个统计程序加以分析，这一程序是以由先前研究所找出来的精算关系为基础所决定的。这些研究结果，毫无例外地显示，统计程序是胜利者。这就表明，精算式预测比临床式预测要准确。事实上，即使是在临床心理医生可以得到比精算方法所用的资料更多的资料的情况下，后者仍然是比前者的预测更准确。也就是说，临床心理医生除

了拥有与精算式预测一样的量化资料以外，还持有与病人个别接触和访谈等个别化的资料，但是这并没有令其预测像精算式预测那么准确："即使拥有信息优势，临床判断仍然不能超越精算方法；实际上，持有更多的信息，一点儿也不能弥补两个方法之间的差距"（Dawes等，1989，p.1670）。产生这种结果的原因当然是统计公式将各种信息数据，以增加预测力为目标作优化整合，而且这一整合是既准确又稳定的。因此，在优化和稳定这两个因素上，精算方法就能远远优于临床心理医生通过非正式方法所收集的资料了。

有关临床式－精算式预测的研究文献中，还包含另一种对二者加以比较的方法，那就是给临床心理医生由精算方式得来的预测结果，让其根据自己与其病人接触的经验，来对这一预测加以调整。结果，在临床医生对精算式预测作出调整后，实际上非但没有增加，反而降低了预测的准确度（见Dawes，1994）。在这里我们又看到了一个不能"接受错误以减少错误"的绝好例子，与前面所述的那个红蓝灯预测实验非常相类。被试不顾灯亮次数多少这一的精算信息，因而不采用每次都预测红灯会亮的策略（可以获得70%的正确率），而是在红灯与蓝灯之间换选，以力求每一次预测都能正确，结果却反而把正确率减掉了12%（只有58%的次数是正确的）。同样的，在这里所述的这些研究中，临床心理医生相信，他们的经验应该可以提供给自己一些"真知卓见"，从而可以使他们作出比定量数据更好的预测。实际上，这些"卓见"根本不存在，他们的预测比依赖公开的、精算的信息所做的预测要差。最后，请注意，精算式预测的优点并不局限于心理学，它也扩展到了许多其他

临床科学中——例如，医学中对心电图的解读（Gawande，1998）。

有鉴于研究表明精算式预测优于临床式预测，米尔（1986）曾说：“社会科学中，没有任何一个争论能像这一量化 / 质化的争论那样，能用从这么大量的、在性质上这么不同的研究领域中所得到的、以及这么一致的及单向的结果。”（pp.373 ~ 374）。但是，令人发窘的是，心理学领域并没有根据这一研究成果身体力行。例如，尽管大量的证据表明，个别面谈这一考试方法实际上是无效的，但在研究生入学与心理健康培训班入班的选拔等方面，还是有许多人在使用这一方法。临床工作者也继续用一些似是而非的理由，来合理化他们为什么要依靠“临床直觉”，而不依赖更有效的总体性预测。例如，戴维斯（Dawes）等曾注意到：

广为流传的反精算论点或误区在于，认为群体统计不适用于预测单个人或事。这种观点滥用了概率的基本原则……。要保持逻辑上的一致，反精算论者必须承认以下这一想法是对的：如果一个人被迫只能玩一次俄国轮盘游戏，但是允许他选择一把镗内装有1发或5发子弹的手枪。由于这一事件的单一性，所以那个人选择哪一把手枪，就变得无关紧要了（Dawes et al., 1989, p.1672）。

这一观点的一个类比是，问你自己对如下的一个科学发现的反应是什么，这个发现是完成过多次类似手术的医生，在下一次类似的手术中成功的概率会比较高（Christensen，1999）。现在有一个医生A，他常做某一手术，失败的可能性很小，而

另一个医生B从没做过这种手术，失败率可能很高，请问你，你愿意让这两个医生中的哪一个来为你做手术呢？如果你相信“概率在单个事件中不适用”，那你就应该不介意让医生B给你做手术。

道斯（Dawes，1994）讨论了对精算式预测的另一反对意见，那就是认为它是“非个人化的”。他指出，其实是个人化或不是个人化是无关紧要的。我们在临床心理学、医学治疗或财经咨询等领域，所争论的是哪一种方法的预测最好。道斯推测说，大多数人会选择一种有效的但非个人化的治疗，而不会去理会一个无效的但是个人化的治疗：

对于感染HIV（艾滋病病毒）的人来说，有什么会比非个人化地去发明一种HIV治疗法更具个人意义及重要性的呢？对于研究生入学申请者、假释犯、心脏病病人，……又有什么会比非个人化的统计预测技术的进步，更具个人意义及重要性呢？关键在于任何一种方法是否能让自己在追求事业的过程中减少失败，能让假释犯在假释期间减少再犯罪，医学治疗更有疗效，等等（p.182）。

承认在诸如预测心理治疗疗效等研究领域里，精算式预测都优于临床式预测，并不会对心理学的声望造成任何损失，因为在如医学、商学、犯罪学、会计学，甚至家畜鉴别等许多内容差别非常大的一些应用领域中，也都发现精算式预测优于临床式预测（见Dawes，1994；Dawes等，1989；Dowie & Elstein，1988）。尽管总体上说，心理学不会因为这些研究结果，而有什么损失，但是对一些个别必须以“专家”身份

来从事临床实务的工作者来说，当然是会造成一定的声誉损失的，特别是那些让病人觉得他们有独特的临床个案知识来帮助他们把病看得更好的从业员，不但声誉会大损，甚至收入也会降低。然而，正如麦佛和瑞特（McFall & Treat, 1999）在一篇论述临床评估之价值的论文中，曾提醒我们的：“我们临床医生所试图评价和预测的事情，本质上是概率性的。这意味着我们不能期望，大自然会如此听话，以致让我们有百分百的把握来预测单一事件。相反地，我们最高的期望也只能是确定一系列可能的结果，然后去估计这些结果出现的相对可能性。从这种概率的角度来看，传统临床评估所内含的那种理想化的目标：要对独特的、不常出现的事件做出准确的预测，其实是一个幻想，反映了我们的无知或自大，或许两者都有”（p.217）。

实际上，如果我们养成“接受错误以减少错误”的习惯，心理学和整个社会都将受益。在试图对每一个不同寻常的单一事件，做出独特的解释时（就我们目前的知识情况来说，独特的解释也许根本不可能），我们常常因此而丧失了对更经常出现的平常事件的预测能力。请大家再一次回想一下红灯—蓝灯的实验，“百分之百红灯策略”令所有出现概率较小，或很少出现的不寻常事件（当蓝灯亮时），所做出的预测都是错误的。如果我们把注意力放在出现概率较小的事件上，而采用“70%红灯、30%蓝灯策略”，结果会怎样呢？我们会在30个不寻常事件中预测对9次（30 × 0.3），其代价是损失了正确预测对21个常见事件的机会，从而把对红灯有70次正确预测的结果，降为只能有49次的正确预测（70 × 0.70）。临床

领域中的行为预测也可以用同样的道理，虚构每一事件的复杂度的解释，可能确实能抓住一小部分不寻常事件，但是必须付的代价是，损失了对大多数事件的正确预测。简单的精算式预测正是对大多数、常出现的事件之预测，做得比较好。加望德（Gawande，1998）指出医学领域也同样需要学习“接受错误以减少错误”这个道理。他认为在医学里，强调直觉、个别化的治疗方法“是有缺陷的——我们试图承认及考虑人的复杂性，非但没有减少错误，反而引来了更多的错误”（p.80）。

道斯（Dawes，1991）曾讲述了一个有趣的故事，在他执教的大学，有人发起了一个罪犯教育项目，该项目包括了一个“与罪犯面谈”的节目。参与这个节目的那个罪犯原来是一个酒吧的配酒员，额外还替人收赌债度日（必要时以枪威逼欠债者）。一天，他被派往俄勒冈州去收一笔钱。当他到达该州，把那个欠债者逼到街角，并掏出手枪要进行威逼时，警察从隐僻处跳了出来，当场逮捕了他。当他向道斯讲完他的故事之后，这个罪犯自信的说，他再也不会入狱了，当道斯问他为什么那么确信自己再也不会被捕时，罪犯回答说，“因为我再也不会去俄勒冈州了！”

这个罪犯的思维有什么错呢？如果我们忽略事情发生的背景，接受他的逻辑，他的结论其实并不坏，正像道斯（1991）所说：

当我讲完这个故事时，人们都大笑起来，但从非概率的意义上来看，罪犯的推论并不是那么坏。他很多次用手枪收债都没有被捕，而他只到过俄勒冈州一次。因而相信一定是俄勒冈

州有什么特别的地方“导致”了他被捕，这是完全合乎归纳推理的原则（p.245）。

此处的问题不在于推断本身，而在于过分集中精力地想去解释一个“特殊”事件。罪犯想解释为什么他会在俄勒冈州被捕——为什么他会在那个特殊的时刻被捕。正是由于这个罪犯过分关注他是在什么特殊情境下被捕，以致使他得出了那个非常荒谬的结论。这个罪犯应该运用概率思维，每一次他用手枪来收债时都有被捕的可能性（也许100次中有3次）。罪犯收了很多次债都没有被捕，这次他被捕了，那只是3次中的一次。一个总体的趋势（他每次用手枪收债就会有“多大多大”的可能性会被捕）解释了他为什么会被捕。这件事发生在俄勒冈州，可能只是刚好，跟本没有什么特别的。

在这个例子中，通过类比看到，忽略较准确的精算信息而依靠临床信息所带来的损害。罪犯就像是一个忽视概率而依靠独有的个案知识来做预测的临床工作者，只不过，他所依靠的是他对自己特殊经历的总结知识来做预测而已。罪犯对其现实世界经历所累积的知识，亦可称为是他个人的洞察力和直觉，使他做出了如果在俄勒冈以外的地方用手枪收债，就不会被捕的临床式预测。相反的，精算式预测会说，如果他继续用手枪收债，在任何地方他最终都会被捕（如果他每次收赌债不被捕与被捕的对比率是9比1的话，那么，他在8次收债中，会被捕的概率已经是超过50%了）。在俄勒冈州发生这一事件也许很特殊，也许不特殊，但道斯（1991）反驳说：

当然，在那被捕事件发生之前，确实有一系列的事件发生，

是在其他事件发生之前没有出现的，但是那是不是就一定要把这一系列事件看作是被捕的“原因”呢？做这样的归因有什么用呢？实际上，试图去寻找这类的原因，非但没有好处，还可能会助长自毁或危害社会行为的延续，因为一旦确定，“为什么”一种行为在某种情境，而不在其他情境造成不良后果，我们就会觉得重复该行为是很安全的，只要避开那个特殊的情境就没问题了（p.245）。

这一收债员的例子，其实与我们真正做的一些研究之结果是非常相似的。确实，有关精算式预测优于临床式预测的研究结果，已经在临床或个案知识充其量最能发挥其价值之处：有关我们自己的知识这一范畴里，也得到了证实。威尔旺，格里芬，林和罗斯（Vallone，Griffin，Lin & Ross，1990）研究了大学生预测在未来一学年中会发生在他们自己身上之事件的能力。被试被要求去预测一个特定事件是否会在他们的生活中发生，并就该事件发生的可能性给出一个确信度的判断。这些事件包括学业上的（例如，今年我将有一门课不及格），友谊上的（例如，我将有一个女朋友/男朋友），与家庭关系相关的（例如，我每个月会给父母打两次以上的电话），和活动上的（例如，我将去旧金山观光，我将加入一个学校的社团）。威尔旺等发现，总体而言，学生对他们的预测过分自信：他们确信自己能预测未来行为的程度，远远比他们应该确信的要大。这一研究重验了一个著名的知识预测实验所发现的一个现象：人们倾向于过分自信（Bornstein & Zickafoose，1999；Brenner，Koehler，Liberman，& Tversky，1996；Griffin &

Tversky, 1992; Schraw & Nietfeld, 1998; Yates, Lee & Bush, 1997)。

当威尔旺等察看了被试在预测自己未来行为时所犯的错误时，发现了一个有趣的现象：这些被试只有在违反了某一特定事件发生的基本概率时，他们才会做出错误的预测。基本概率是指把所有参与的被试看作是一个整体，用某一特定事件在他们身上发生的总人数来计算该事件发生的概率（发生人数除以总样本人数）。有一些事件发生的基本概率比较高（一个样本中大多数人都会发生那些事件），另一些事件发生的基本概率比较低（一个样本中很少人发生）。几乎所有学生犯的错误以及所有他们显示的、错误的过分信心，都是在违反了这一基本概率的情况下出现的。违反基本概率常会表现在被试说出这样的话来：“啊，这虽然可能会发生在大多数人身上，但不会发生在我身上。”或者“这虽然在大多数人身上可能都不会发生，但它会发生在我身上。”实际上，如果被试以基本概率来预测自己的行为，而不是把自己看作是一个例外，他们会预测得更准确一些。由于认为自己独具有关自己的知识，因此深信对自己行为的预测会更准一些，结果，这种自信反而使他们的预测更不准了。

这一研究又是一个临床式—精算式预测的例子，而且也是一个不遵循“接受错误以减少错误”的例子。正如威尔旺等（1990）所指出的那样，“有关行动者的知识不管有多少，即便是包括有关行动者自己及其过去经历的知识，也无法让预测者，像明智地运用精算的基本概率知识那样成功地提高其预测的正确性和改进其精密度”（p.590）。

华格纳和科瑞（Wagenaar & Keren，1986）阐述了，对个人自我知识的过分自信以及对统计信息的大打折扣，是怎么样地把宣传系安全带驾车的交通安全推广活动打垮的。这都是因为人们总是认为："我和别人不一样，我驾车很安全"问题是85%的人都认为"自己的技术比一般驾车者高明"（Svenson，1981）——这显然是很荒谬的。

相信"统计数字不适用于单一个案"所造成的类似之谬误，是致使赌徒恶习难改的重要因素。华格纳（1988）在他的赌博行为研究中总结道：

> 从我们和赌徒的讨论中，可以非常清楚的看到，赌徒大体上都能意识到赌博的长期后果是不利于自己的。他们也知道最终输的会比赢的多，而且永远都会是这样。但他们却不能把这些总体的统计考虑应用到下一局、下一小时或下一个晚上，满筐满篓的秘籍、贴士、灵感等让他们觉得，统计学在下一局或下一小时里是派不上用场的。而他们相信自己还是可以预测下一局的结果的（p.117）。

华格纳发现，强迫性赌徒有很强烈的倾向不去"接受错误以减少错误"。例如，二十一点牌局的玩家，普遍存在拒绝使用一种叫基本率的策略（见 Wagenaar，1998，第二章）的现象，基本率策略是一种保证可以把庄家的赢率从6%或8%减少到不足1%的策略。基本率策略是一个长期性的统计策略，必须在一段长时间使用的情况下才会见效。强迫性赌徒之所以拒绝不用，是因为他们相信"有效的策略应该是在每一次都有效才对"（p.110）。在华格纳研究中的赌徒"总是说，像基本率这样的统

计系统所教授的一般性策略是不会有用的，因为它们忽略了每一具体情境的独特性”（p.110）。抛弃能保证他们少输上千美元的精算式策略不用，这些赌徒徒劳地转去追求那些建立在每一具体情境之独特性基础上的临床式预测。

当然，在这里有关临床式—精算式预测研究文献的讨论，并不表示个案研究在心理学中没有作用，请大家记住，这一章所谈的，只是预测行为这一特殊情境。回想一下在第四章对个案研究之价值的讨论已经指出，个案信息在引发对重要的、需要进一步探研之变量的关注方面是非常有用的。而这一章中所说的则是，一旦重要的相关变量已经确定，而我们要开始运用它们来预测行为时，测量这些变量，并使用统计公式来进行预测，则是最好的研究方法及程序。首先，通过精算式方法可以得到较准确的预测；其次，用精算式程序所得出的预测是一种面向公众的公开知识，任何人都可以使用、修改、批评或争论。相反地，使用临床式预测就等于要依靠个别权威的个人学历及经验，而他的评估正因为其声称是单一及个人独特的，因此也是不可能付诸公众的批评与监督的。正如同道斯所说的：

一个专业心理学者会在法庭上声称：一个极具杀伤力的“临床判断”是基于他自己多年所累积的经验，是无法公开被证实的、只能在诸如专业资历、给予相似的证词有几年、态度品行等不相干的地方接受挑战。相反的，一个统计公式或模型是面向公众的，所以它可以随时在理性的基础上接受挑战（p.104）。

小　结

偶然在心理学中的作用常被外行的大众和临床心理从业者所误解。人们很难认识到，行为事件结果的变化有一部分是由偶然因素造成的。也就是说，行为的变化有一部分是随机因素作用的结果。因此心理学家不应该以自己能预测每一个个案所涉及的个别行为而自诩。心理学的预测应该是概率性的，是对总体大趋势的概率性预测。

让人觉得自己可以在个体层次上进行心理预测，是临床心理学家常犯的错误。他们有时候会误导他人认为，临床学历及经验赋予了他们一种对个别案例能做准确预测的"直觉"能力。恰恰相反，几十年来，值得一看的研究都一致性地表明：在解释人类行为的原因方面，精算式预测（基于群体统计趋势的预测）远远优于临床式预测。目前还没有证据表明，临床直觉能预测一个统计趋势是否会在一个具体特定的个案中出现。因此，当一个人要预测行为时，千万不要把统计资料搁在一边。统计预测也正确地告诉我们，当一个人在预测人类行为时，错误和不确定性是永远不可避免的。

（韦庆旺　译）

第十二章　让人瞧不起的心理学

虽然公众对心理方面的话题总是带有浓厚的好奇心及神秘感，但是他们对这一领域及其所取得之成就的评价，却多半是负面的。心理学家们都意识到这个“形象问题”，但他们大多又感到束手无策，于是他们只能不去理会它，假装这个问题不存在。然而，这样做是一个错误。当大众传媒在形塑公众对事物的感知上变得愈来愈具影响力时（例如，小说式的电视纪实剧里所说的，对那些不读书不看报的一般老百姓来说，就已经是真正的历史了），不理会心理学的形象问题，只会使这一问题越变越糟。

罗德尼·丹杰费尔德（Rodney Dangerfield）是美国一个广受欢迎的喜剧演员，他有一句抱怨的话现在已经成为他的标志：“我老是被人瞧不起！”从某种意义上说，这也正是心理学在一般老百姓心中的地位。本章就是想谈谈为什么心理学会像丹杰费尔德那样无法得到应得的尊敬。

心理学的形象问题

有关心理学形象问题的成因，有些在前面已经讨论过了。举例来说，在第一章中所讨论过的弗洛伊德问题就是其中之

一。弗洛伊德理论的不能被证伪性，无疑促成了人们对心理学的低评价。如果公众认识任何一个心理学家的话，这个人不是弗洛伊德就是斯金纳。对他们理论的歪曲报道及传播，一定会导致公众认为心理学是一门非常肤浅的学问。当一门学科中最有影响力的学者被认为是主张人是没有思想的，以及人和老鼠没有差别的，那时这门学科在外界公众心目中，还有什么尊严可言？当然，斯金纳并没有否认人类能够思考，他从动物身上发现的许多有关工具性条件反射的定理，已被证实的确是可以推广到人类行为上的。然而，公众对这些科学事实却知之甚少。同样地，对弗洛伊德理论的歪曲也降低了心理学在公众心目中的形象。

心理学和超自然心理学

除了弗洛伊德和斯金纳的研究之外，一般人对其他卓越心理学的研究几乎一无所知。为证实这一点，你可以到附近的书店去看看公众在那里可以购买到哪一类心理学的读物。调查结果你会发现，那些摆在书店卖的心理学读物通常可以分为三类。第一类是一些心理学的早期经典著作（弗洛伊德、斯金纳、荣格、弗洛姆、埃里克森等），而这些著作几乎已和当代心理学毫无瓜葛了。

在多数书店中可以找到的第二类读物，是那些假冒为心理学的“伪科学”书籍，里面充斥着永无止尽的一连串有关超常现象的记载及报告，如心灵感应、千里眼、意念移物、超前感知、转世、生物规律、星灵投射、金字塔力量、植物沟通、通灵手术等（Lilienfeld，Lohr，& Moirer，2001）。

书店中这类所谓"心理"读物的大量存在，无疑导致并同时反映了人们对科学心理学的误解。他们认为心理学家就是去证实这些超常现象存在的人。这种误解对心理学而言无疑是一个讽刺。事实上，心理学与那些超常现象是毫无关系的。现代心理学有兴趣研究的东西完全不包括那些超常现象，其中缘由很可能会让许多人大吃一惊。

要说对超感知和其他超常能力的研究不能被认为是心理学的一部分，可能会引起许多读者的不满。多个调查结果一致显示，超过50%的公众相信有超感知现象的存在，并且对这一想法坚信不疑（美国国家科学委员会，2001）。历史研究和调查可以让我们对公众之所以会如此热衷于这些信念的原因看出一些端倪（Alcock，1987； Bainbridge & Stark，1980； Grimmer，1992； Stanovich，1989，2004）。物质至上的文化削弱了许多人的传统宗教信仰，这些人就开始寻找其他形式的神秘主义来填补这一空档。像许多宗教一样，许多所谓的超常现象也标傍像转世之类的说法。对部分人来说，这一来世的说法能满足他们想要超越现有生命极限的需求。这就可以说明为什么，当心理学家出来说，用科学心理学研究方法无法证实超感知现象的存在时，人们感到无法接受。声称心理学不把超感知视为一个可行的研究领域，总是让那些信徒们感到生气，他们指责心理学家把这一类的研究课题摒除于心理学研究之外是太武断了。当心理学家对这些人提出的抗议采取不理不采的态度时，他们对公众对心理学的误解，完全没有起到减轻的作用。心理学家的确应该针对为什么这些抗议是没有根据的，给予一个谨慎而清晰的解释才对。下面就是我们

对这一误解做出的解释。

科学家们并不是根据什么经文来确定研究课题的。也没有什么文稿指出什么是可以，什么是不可以被研究的。研究领域的出现、扩展或终止是依据理论与方法的自然淘汰规律。也就是说，那些能够产出硕果的理论和实征发现的领域，就会吸引大量的科学家继续研究。而那些陷入死胡同的理论，或者那些没有能被重复验证的实征领域就会被摒弃。这种对理论与方法的自然淘汰过程引导了科学不断地接近真理。

例如，超感知在当代心理学不被认为是一个可行的研究课题，就是因为对它的研究一直无法累积任何正向的成果。因此，心理学家就不再对它感兴趣了。在这里我要强调“当代”这一词，是因为心理学家在早前确实曾有一段时间，对超感知领域是怀着极大的兴趣的，只是在累积了大量的负面证据之后，才对它失去了兴趣。正如历史所展示的那样，研究课题通常不是由政府某个权威机构宣布停止研究的，它们只是在生存竞争的环境中被自然淘汰出局了而已。

在心理学领域里，超感知从来没有被认为是一个不能研究的课题。有关这一说法的证据是清楚并且很容易取得的（Alcock，1990；Druckman & Swets，1988；Hooft，2000；Hyman，1996；Marks，2001；Milton & Wiseman，1999）。过去有许多研究超感知的论文发表在正式的心理学研究刊物上。那些在媒体上频频曝光的超自然心理学家总喜欢让人们觉得这一领域是崭新的，借此以为惊人骇世的新发现问世了。实际上，事情根本就没有那么戏剧性。

对超感知的研究，可以说，是和当代心理学的历史一样

的久远，它并不是一个新的研究领域。在心理学文献中，它也曾经像许多现在被认为是可行的课题一样的被认真地研究过。然而，在正式心理学杂志上所发表的、有关这一领域的许多研究结果，大多显示了它是子虚乌有的。在被研究超过了90年之后，我们仍然无法在严格控制的情境下，成功地重验任何超感知现象。尽管过去几十年来做了大量有关超感知的研究，却从来没有一个研究可以达到这一简单的基本科学标准。甚至许多超自然心理学家和其信徒自己也都不得不承认这一点（见Alcock，1981，1990；Druckman & Swets，1988；Krippner，1977）。也正是因为这个原因，才令当代心理学家对这个课题失去了兴趣。

有时人们会被误导认为，科学家否认超感知现象的存在是因为它违反了当代已被普遍接受的、有关大自然的理论。从第一章和第二章所讨论的科学程序来看，很明显地，这种说法是不对的。所有的科学家都忙于推翻现在已被接受的、有关自然事物的理论，因为，只有在不断地通过改变和精化现有的观点，当然还要在遵循“关联原则”的大前题下，才能逐渐接近真理。当一种新的现象与现有的知识结构相冲突时，科学家们会对此提出疑问并寻求其他可能的解释。所以，心理学家不相信有超感知现象，并不是因为它与原有的知识累积相矛盾，而是因为根本没有任何科学的证据来支持它的存在。简而言之，没有被证实的现象自然不需要作任何科学的解释（见Alcock，1981，1984，1990；Hines，1988； Humphrey，1996；Hyman，1992，1996；Milton & Wiseman，1999）。

现在，颇具讽刺的事来了，心理学家虽然找不到这类证

据，但是在许多测验超常能力是否存在的大会上，却都扮演了非常重要的角色。他们的重要性可能仅次于那些专业的魔术师，这些魔术师当然是下了最多功夫去拆穿多数超常能力演示其实都是骗局的功臣（Randi，1986，1987）。而且，很多相关论述，以及质疑有超常能力的重要书籍，都是出自心理学家之手。

这个现象的讽刺性是很显而易见的。心理学作为一门可能是目前对超感知的存在问题最能给予精确评估的学科，在公众的心目中，竟然正是被与这些伪科学混为一谈、两者难分难解的学问！正如下面还要再详细讨论的，心理学常常会陷入这样一个“里外不是人”的双重灾难之中，这里所说的仅仅是其中一个例子罢了。那种认为在心理学里没有什么行规的想法，以及认为这个领域缺乏科学标准来判定其知识之可信度的信念，常是导致人们把心理学与超感知这样的伪科学联系在一起。然而，如果心理学家有朝一日真的能够成功地让公众认识到这些伪科学的真实面目，那么，又由于在人们心目中心理学与伪科学总是联系在一起，这些证据可能被视为是说明“心理学不是一门科学”的铁证！你看，有多倒霉！

自助读物

在书店里常见的第三类心理学读物就是所谓的自助读物。当然，这类读物也有许多不同种类。（见Fried，1994，1998；Fried & Schultis，1995；Paul，2001；Santrock，Minnett， & Campell，1994）有一些书是励志型的，目的是为了要提高人们的自我价值感和自信心。另一些书试图将一些

熟悉的老生常谈用新的表达方式包装起来。只有少数（实在是太少了）的几本是由有责任心的心理学家为公众所撰写的。还有许多书，不包括在上述第二类中，为了确保自己的“独特”性而声称自己发明了一些新的“疗法”，不但可以治愈某些特殊行为问题，而且还能满足老百姓的一般性的需求（赚多一些钱，减肥和提高性生活质量是其中的“三大”），从而得以销量大增。这些所谓的新疗法的成效很少建立在控制实验研究的基础上。如果作者是个临床医生的话，他们通常只是依靠他们的个人经验或者少数的几个病例，来支持他们的新发明。

许多行为和认知疗法，都曾通过相当严格的心理学检验程序来支持其有效性，但是有关这些疗法的书籍却很少出现在书店的书架上。电子传媒的情况就更糟糕了，电台和电视台几乎没有任何正儿八经的心理学报道。相反地，他们总是请来许多伪“疗法”的散播者和爱出风头的媒体名人来上它们的节目，而这些人与心理学这个领域可谓是风马牛不相及。媒体之所以会这样做的主要原因是，正规的心理治疗法从来都不会声称自己能令病人马上痊愈，甚至不会担保治疗一定成功，或者夸大其治疗的范围（例如，“你不仅会把烟戒掉，而且你生活的方方面面都会得到改善！”）。同样地，现在互联网上也出现了类似的情形——缺乏严格的同行评审，在网上推销的治疗方法通常都是不可靠的。

公众在很大程度上并不知道，出版业并没有为读者就自助性的读物把关，提供质量保证。弗里德森（Fridson，1993）曾讲述了一位大众心理学读物作家的故事。这位作家写了许多

有关就业方面的畅销书。他在《财富》杂志上有一个专栏，并且经常在一些电视访谈节目上出现。不久，人们发现这个作家并没有像他本人声称的那样拥有博士学位（事实上，他得到的是一个未被政府认可的函授学校的博士文凭）。进一步的调查揭示，这位作家声称自己曾为研究一个问题所做过的8000次访谈中，有一些是自己编造出来的。当人们质疑关于出售这些骗子所写的、具误导性的书籍背后之商业道德问题时，这位作家的一个出版商说，“我们出版商做的事有99%是基于对作者的信任”（p. 144）。阅读本书的读者，你们从今以后要对“信任”一词附上一个重要的辅助性脚注：要具备有关现代科学验证方法背后之逻辑的知识。

在美国书市中占相当比重的这些自助性读物，使人们对心理学的一般看法产生了许多不良影响。首先，像弗洛伊德的影响那样，这些书使公众搞不清心理学研究所关注的焦点在哪里。举个例子来说，虽然有相当数量的心理学家在为肥胖、人际关系和性问题提供治疗，并且在不断地做研究，但这个数量比起自助读物中所说的要少得多。这种误解也使得公众以为大多数的心理学家都忙于关注异常行为的研究和治疗。事实上，大部分的心理学研究的是人类的正常行为。

除了引起对研究内容的误解之外，自助性读物还让人们对心理学的研究方法和目的产生错误印象。正如第四章中所讲的那样，心理科学并不认为几个个案研究、见证叙述和个人经历陈述就能构成充分的实征证据，用以支持某种疗法的有效性。然而，这些方法却正是大多数自助“疗法”所提供的支持证据之类型。自助类的读物因此误导了社会大众，使他

们认为大多数的心理学理论就是在这一类证据的基础上得出的。在第八章中已阐明，一个理论的验证需要许多不同类型的证据来支持，其中个案研究证据所提供的支持是最弱的。视这一类证据为证实某一理论或疗法的确凿证据，是犯了根本性的错误。

“食谱”知识

最后，自助类的读物使公众混淆了心理学的目标和多数心理学研究所追寻的知识类型。心理学家利 · 夏法（Leigh Shaffer）认为这种读物留给人们一个深刻的印象，认为心理学研究者所追求的是那种“食谱”式的知识（Shaffer, 1981）。食谱式的知识是指那些只知道如何去使用某物，但对其背后运作的基本原理一无所知的、“知其然，而不知其所然”的知识。例如，大多数人知道如何使用电话的知识，他们知道如何拨号，如何获得信息，如何打长途等等。但许多人对电话运作背后的物理原理一无所知。他们不知道电话是怎么样实现通话的功能，他们只知道他们能够使用它。这就是电话的“食谱知识”。在我们的社会里，许多有关技术性产品的知识都是食谱式的知识。

当然，这也不完全是一件坏事。事实上，多数技术产品就是按照用户对其背后的运作原理一无所知的条件来设计的。食谱知识这一概念提供了一条可以区别基础研究和应用研究的途径。基础研究工作者寻找自然界的基本原理，而不去考虑这些原理能否转化为食谱知识。应用研究工作者则致力于将基本原理转化为一些实用的产品，这种产品只需要食谱知识就可

以使用了。

多数自助读物只提供关于人类行为的食谱知识，它通常是以下面这样的形式来呈现的，“你只要做X，你就会变得更加Y了”，或者“做A，其他人会对你更B”。当然，如果这个食谱是正确的，做这些事也不一定就是错的，许多正规的心理治疗都提供了大量食谱知识。然而，当人们错误地认为，所有心理学研究的终极目标就是提供食谱知识时，问题就产生了。虽然很多心理学研究者确实致力于将基本的行为理论转化为可以应用的心理治疗法、保健行动方案或有效的工业组织模型，但心理学主要还是一门发现行为的普遍事实和理论的基础学科。在这里，我们看到心理学研究为何会被外界看为是很“怪”的另一个原因：基础理论的研究与应用研究有很大的差异，

如果有一个人走到分子生物学实验室，询问一位研究者，我们在头痛时是应当服用两片还是3片阿司匹林，我们会觉得这个人很无聊。原因并不在于分子生物学与止痛药没有任何关系，将来对止痛药的研究可能会运用到这一领域的知识。这个人问的问题很无聊，是因为分子生物学家并不是那种开药方的人，他不能解答你是要吃两片还是3片阿司匹林的问题。研究者所关注的是有关生物物质在分子层面的基本数据。这些数据可能会为许多不同领域提供食谱知识，但发现基本数据和将这些数据转化为食谱知识的人不大可能是同一个人，转化为食谱知识的方法，也会与最初发现事实的方法有所不同。

由于自助读物导致公众相信多数心理学家是在追寻食谱知识，心理学家们实际所做的大量基础研究，看起来就有些

"怪怪"的了。亨奇特（Hecht）曾让被试在一间黑暗的房间里注视着一个小红灯，到底与我们的现实世界有什么关系？是的，从表面上看来，确实没有一点儿关系。亨奇特是对人们的视觉系统如何适应黑暗的基本原理感兴趣，这些基本原理最终会转化成食谱知识，告诉人们如何处理一些特定的问题，例如因维生素缺乏而导致的夜盲症。然而，这一转化并不是由亨奇特本人来完成的，而且它要在几年之后才会到来。

因此，自助读物给公众对心理学的感知带来了两个不良的副作用。第一，这些读物里谈论的问题并不一定是当代心理学关注的焦点，它反而是反映了消费者想阅读的内容。心理学系的学生通常无法充分地意识到，出版是一个商业行为，市场的力量决定了什么样的读物应当搁在书店的书架上。然而，科学的关注并不是这样决定的。在所有的学科领域，尤其是在心理学，科学家们认为有生命力的念头与那些被商业包装起来的具有销售力的念头，完全是两码事。

第二，自助读物让心理学看起来完全是一种食谱知识。这是完全错误的，它忽略了心理学领域所进行的大量基础研究。

心理学和大众传媒

大多数心理学家本身已经意识到他们这一学科有公众形象问题，同时，这一事实也导致了一些不良后果。在这里，我们充分看到这一领域"祸不单行"的情况。许多心理学家非常关注要如何将他们的学科，在公众的脑海中与自助读物和伪科学划清界线这一问题。这种关注使许多心理学家小心翼翼，

避免对人类迫切需要解决的问题提出肯定的解决办法，以免自己与那些自助读物作家或伪科学散播者“同流合污”。这种不愿承认自己在某一社会关切的问题上有专门知识的做法，已经变成各大学的研究生院训练心理学研究者的基本态度了。心理学研究已经是最具自我批评力度的学科之一。心理学杂志充斥着各式各样的论文针对各种研究方法的问题进行批评及警告，有些问题在前面几章中都有提到（例如，从相关关系推断到因果关系的陷阱，集中证据的原则，以及检验其他可能解释的重要性等等）。

这样一来，心理学向外界公布其研究结果时，所采取的保守做法，就根植于大多数心理学研究者的心中。心理学家也因此不愿意对外声称，他们对于目前人类迫切需要解决的社会问题有任何决定性的答案。这样做，当然，往往是对的。人类行为的问题是错综复杂的，而且要研究起来并不容易。谨慎一点儿是好的。然而，当心理学与媒体打交道时，就不是那么一回事了。这一保守做法产生了一些不幸的后果。

大众传媒的独特逻辑是，如果公众对一个特定的心理问题感兴趣，媒体就会去“制作”一个故事，而不管是不是真有一个故事可讲。当一位科学家告诉一位记者：“对不起，那是个复杂的问题，还没有相关的研究资料，所以我不想对此发表意见”，这样的回答绝对阻止不了那位记者去别的地方，继续寻找答案，直到他/她找到了为止。他/她所找到的，通常是一位不那么保守的科学家（对心理学问题而言，他/她通常只要找到一个能够被称得上是“权威”的人即可），这位人士就一定不会羞于下结论了（Murray，

Scwartz, & Lichter, 2001; Ruscio, 2000)。

柯力、偌顿和卡维尔（Kelly, Rotton & Culver, 1985）讨论了这一非常灾难性的“媒体逻辑”如何使“月亮效应”变成家喻户晓。所谓月亮效应，就是指相信月亮的盈亏会影响人类行为，特别是异常行为。柯力等人分析了37个不同研究的结果，发现不存在人类行为受月亮效应影响的证据。反而，他们发现 “媒体效应”的影响很大，媒体效应就是指媒体能够指鹿为马，使人们相信一些根本不存在的现象。柯力等人不客气地挑明了我们刚才所讨论的媒体逻辑：

报纸、电视节目和电台的脱口秀总是偏爱那些宣称满月能影响人的行为……，当记者打电话给我们，如果我们确认他们的想法，说“当满月时，街道上到处都是失去理智的人”，他们自然会高兴得多。不幸的是，我们不能那样做。当一个科学家不能给他们一段能够让他们转化为煽情标题的言论时，他们总能找到一个可以提供他们所需言论的所谓“专家”（p.133）。

新闻学教授柯蒂斯・麦克道格尔（Curtis MacDougall）在他的著作《迷信和报业》（1983）中，给出另一个媒体逻辑的例子。有一位记者经常写通灵能力方面的故事，当问及他本人是否相信真有通灵能力这回事时，这位记者回答：“我没有必要要相信它。我所要的是两位有博士学位的人告诉我，事情是如何如何发生的，我就可以写一篇报道了。”（p.558） 这个例子阐明了为什么在本书中，我们一再强调，经过严格同行评审的科学杂志里给出的信息，才是可靠的的原因。这类杂志

是惟一能够为读者提供科学信息的地方。电视脱口秀没有对信息进行科学的评审，他们正是按麦克道格尔所说的逻辑行事：如果有人在说一件事，并且公众对这件事感兴趣，那就把它搬上屏幕。测试、检验、真相、证据、逻辑、辩护、数据等等，这些与脱口秀没有任何关系，脱口秀只是提供某一时段的娱乐而已。要找科学信息的话，请去别的地方找吧。

媒体将科学传播给公众时的选择过程，是基于以下的逻辑：那些言辞谨慎的科学家很少会讲出什么值得重复的名言，从而引起公众的反响，只有那些大胆、愿意冒险的人才会成为公众人物。当然，这也不见得总是一件坏事。举个例子来说，已故天文学家和电视名人卡尔·萨根（Carl Sagan）有时会比他那些保守的同事做出稍为大胆一些的推测（Poundstone，1999），但是他在普及天文学方面的巨大贡献无疑超越了他个人言论的一些不正确的瑕疵。

心理学的情形却完全不同。多数在媒介露脸的所谓心理学家，不是萨根类的人物，他们在自己的研究领域的同行眼里，绝对是没有任何地位的。心理学之所以如此不同，是一系列因素综合造成的。比起自然界的其他方面，人们更迫切地希望找到有关人类自身问题的答案。比起土星带的构成、宇宙是否可能真有黑洞，人们更想知道如何减肥，心理治疗是否有效，小别是否真的更胜新婚，如何提高小孩的学业成绩。这一解决自身问题的紧迫性，再加上人类问题的复杂性，致使要寻找它们的答案比较困难。这就让我们比较清楚地看到，为什么有责任心的心理学家会比其他科学家更不愿意在媒体上发言了。因此，在其他的学科领域，媒体的选择过

程帮他们铲除的是那些保守的科学家，取而代之的是那些对下结论要求不是那么严谨的科学家。不幸的是，在心理学这个领域里，媒体的选择将所有的科学家都铲除出局了！

心理学家在面对媒体记者时，为维护科学尊严所把持的保守态度，往往制造了一个媒体空白，因为不确定的、有保留的说话是编造不出什么有趣的故事来的。不管合理不合理，这个空白不久就被填补起来了。许多自助大师和通灵骗子乘虚而入，并且在电视和电台的脱口秀节目里炒得非常红火，在公众的心目中这些伎俩逐渐与心理学联系起来了。简言之，保守主义的后果是适得其反。本来心理学家按照科学的谨慎原则向公众发表研究结果，现在这种保守主义的做法反而使得伪心理学在媒体大行其道，从而减少了公众和其他科学家对心理学的尊敬（见 Pallak & Kilburg，1986）。

美国心理学会会刊（Monitor）上的一篇文章（Azar，1998）统计了《纽约时报》在报道自然科学相关的文章时，引用科学参考文献的比率为 51.4%，而在社会科学相关的报道中只有 14.3% 引用了科学参考文献。无独有偶的，心理学家鲍伯 · 切尔迪尼（Bob Cialdini）引述了一项研究，发现在报纸和新闻广播中，当讲到自然科学研究的作者时，有 80% 的时候会称他们为科学家，而对社会科学研究的作者，则只有 20% 的时候会被这样称呼（Cialdini，1997）。相反地，这些社会科学研究的作者，有 80% 的时候是被称为是作家或作者。这就大大削弱了社会科学研究的科学性。

有时，在报道有关心理学的话题时，媒体甚至懒得假装曾调查过一个报道的科学真实性，仅仅只是在嘴皮上说说要注

意这一问题，就算过去了。我们来看看1994年6月20号《新闻周刊》上的一篇文章。在用了一条长栏宣传一种新的疗法之后，《新闻周刊》引述说，其他新闻杂志把这一“眼动去敏和再加工”治疗法称为是“奇迹”，并写道：“不幸的是，对这一疗法的热烈反应，主要是建立在见证证据的基础上。（见第四章）支持者们还需通过科学的方法验证，来显示这种疗法确实比其他形式的治疗更具独特优点”（Cowley & Biddle，1994，p.70）。那么，你可能会问，既然还没有足够的科学证据证明这一疗法是有效的，那发表这篇报道的用意何在？为了要强调媒体问题给心理学造成多么坏的影响，我们只需要去看看这篇文章中所介绍这一疗法的信息来源即可。《新闻周刊》可不是那些超市小报。你也许会诘问：将一个没有被验证过的疗法发表在一个全国性的周刊上，这背后到底可能隐藏着什么样的目的呢？这本周刊将会被成千上万的人读到，他们之中有许多人可能正在拼命地寻找可以治愈各种不同疾病的治疗方法。在这篇文章发表两年后，心理学家史考特·利利恩费德（Scott Lilienfeld）评估了有关这一疗法的疗效证据，发现大多数的支持证据是来自那些未受控制的个案报告（见第四章）。设有控制组的实验研究则表明没有任何证据显示这个疗法是有效的（Lilienfeld，1996）。

心理学和其他学科

当然，心理学并没垄断对人类行为的研究。许多其他的相关学科，采用不同的技术和理论视角，也对心理学有所贡

献。许多关于行为的问题都要求采用交叉学科的方法来研究。然而，在这样做的同时，大多数心理学家必须要接受一个非常残酷的事实，那就是，当这种学科整合的成果发表时，心理学家的贡献往往会被其他学科所掩盖。

关于心理学家的贡献被忽略、减小，或者吞占的例子不胜枚举。例如，第一个有关电视暴力对儿童行为影响的研究，是由美国外科医生总署主持的，研究结果发现二者之间存在因果关系，之后由美国医疗联会通过一项决议，来重新确认该项研究结果，并且去广为宣传。当然，这一点儿也不令人惊奇，医学单位主持的研究由医学组织来接受认可。但是这一举措在无意间所造成的一个后果是，媒体不断地将电视暴力的研究成果与美国医疗联会联系在一起，使公众误以为这个研究都是由医学专业人士做的。事实上，绝大多数有关电视暴力对儿童行为影响的研究都是由心理学家完成的。

心理学家的工作经常被划入其他学科的另一个原因是，随着时间的变迁，心理学家一词的含义已经变得含混模糊了。许多心理学研究者通常会在自己的称谓上把自己的研究专长添加上去，例如生理心理学家、认知心理学家、工业心理学家或者神经心理学家。还有一些人甚至用一些称谓是跟心理学毫无关系的，例如神经科学家、认知科学家、社会生物学家、人工智能专家和行为研究者等。所有的这些举动，再加上媒体认为“心理学不是一门科学”的偏见，都导致了心理学家的成就被误划入其他学科：生理心理学家的成果被划入生物学，认知心理学家的被归入计算机科学，工业心理学家的被归入工程学和商学等等。可悲的是，就连有一位杰出的心理

学研究者，卡尼曼，获得了诺贝尔奖，心理学的贡献也不会被人注意到，因为他获得的是2002年的诺贝尔经济学奖！诺贝尔的奖项中没有设心理学这一项，这也充分说明心理学的贡献不为人所认可（MacCoun，2002）。

埃莫瑞大学耶克斯灵长类动物研究中心的负责人，心理学家弗雷德里克·金（Frederick King）告诉我们这样一个故事。金在羊角疯问题的神经和行为研究方面经验丰富，有一天当他花了大量时间向一位记者解释，动物模型对人类神经失调研究的重要性后，这位记者问道：“你怎么会知道这么多关于羊角疯的数据的？你不是一个心理学家吗？”（King，1993）。

在20世纪70年代末，法院判决了几起有关标准测验的诉讼案。其中的一起，是帕西起诉汉农的案子，涉及智力测验中的文化偏见问题。审理该案的法官认为能够帮助他裁定的惟一方法是，自己亲自检查每一道测题，并且根据自己的直觉判断来决定这些测题是否真的有文化偏见。他对自己能做出正确判决的能力深信不疑，并对测验的每个问题写下自己个人的看法（Bersoff，1981；1982）。这位法官断定，在这些标准测验中，有一组测验中有8个题目，而另一组有一个题目可能是带有偏见的。这个法官没有意识到，像这样的问题其实可以通过科学的验证方法来解答。个人意见不仅与真相无关，甚至可能是极具误导性的。决定这些标准测验题目是否带有偏见，涉及复杂的统计程序和大量的数据收集。心理学家在为这些评估工作，进行必要的数据收集及发展相应的统计技术等方面都做出很大贡献。

颇具讽刺意味的是，对这一问题所做的实际实验研究结果发现，普通人对于这些题目具有文化偏见的直觉判断通常都是错误的。许多被认为是没有偏见的题目实际上却含有不同的种族和社会文化歧视，而许多在表面上看来是不公正的题目统计结果却没有显示出有偏见（Sandoval & Miille，1980）。举个例子来说，韦氏成人智力量表在加拿大受到批评，因为其中某一分量表（“信息”分量表）的一些测题看起来好像偏袒美国公民，让他们得分会比较高。例如，其中有一个题目要求被试写出 1950 年之后美国 4 位总统的名字。于是，当这个测验在被搬到加拿大来实施时，这一题目就被“加拿大化”为“1900 年之后加拿大 4 位总理的名字”。然而，就算在这样一个非常不起眼的、“常识性”的改变上，也产生了一个小问题：加拿大公民在“总统版本”的智力测验得分，反而要比在“总理版本”的得分要高。

我们是自己最坏的敌人

为了避免看起来好像在心理学的形象问题上，我们只会责怪别人，现在是我们来检讨心理学家自己的时候了。试图把真正的心理学向公众介绍的正规心理学家往往得不到什么好的回报。然而，美国心理学会（APA）和美国心理协会（APS）现在都比以前更加努力在促进与公众的沟通，美国心理协会还为了这一目的开办了一个新的期刊：公众利益下的心理科学（Psychological Science in the Public Interest）。心理学需要在这一方面再加把劲。不然，将来如果公众还对这一学科心存

误解时，我们就只能说我们是自作自受了。

美国心理学会前主席罗纳德·福克斯（Ronald Fox）在最近的致辞中提到了心理学的一些沟通问题以及我们自己是如何导致这些问题的产生：

一些经常在大众传媒上露脸的执业者，他们的做法是不专业的，不道德的，并且会让他的同行蒙羞的……我们的学科对那些不负责任的、令人发指的公开欺骗，缺乏有效的对策……在今日的世界里，公众经常被媒体喂食一些骗子的观点和意见（在最近的一个电视脱口秀节目里，有一个心理学家声称他已经帮助许多病人回忆起上一世所受的精神创伤），而把那些明智执业者的观点和意见弃之不用（Fox，1996，pp. 779～780）。

最后，心理学内部也存在部分反科学的态度（Coan，1997；Watters & Ofshe，1999）。例如，在一些心理治疗的圈子里，他们一直都拒绝对自己所用的疗法进行科学的评估。专栏作家和心理治疗医师查尔斯·柯瓦斯阿默（Charles Krauthammer）写了一篇文章，论述了这种态度对心理治疗的诚信度所造成的严重损害（Krauthammer，1985）。第一，由于没有机制把那些无效的疗法排除掉，造成各种疗法泛滥成灾。这种泛滥不仅使消费者的权益受到损害，而且还造成这一领域的混乱："心理治疗已经进入混乱的状态，因为……它容许太少的流派消亡，它无法在内部实现有效的淘汰，心理治疗正在鱼龙混杂中走向死亡。" 柯瓦斯阿默在这里所悲叹的是，不能遵循证伪原则已经妨碍了这一领域的科学进步。

第二，柯瓦斯阿默又指出了心理治疗这个圈子的一个内在矛盾，一方面，他们认为心理治疗更像一门艺术，而不是科学，因此反对采用科学的方法来评估它，另一方面，他们仍然非常关注政府和健康保险公司会不会让接受他们服务的消费者报销。柯瓦斯阿默揭示了这两种态度的内在矛盾："只要心理治疗业拒绝为他们的工作提供科学证据，他们就会陷入经济困境。毕竟，如果心理治疗真的是一门艺术的话，他们应当由国家人文基金提供资助，而不是医疗机构。"与这种观点相呼应的是，霍华德（Howard）等人（Kopta、Lueger、Saunders & Howard, 1999）在他们对心理治疗效果研究的评估中写道："某种特定疗法的效果必须通过实征方法来检验其有效性，才算是有足够的理由去取得向保险公司和有管理的护理公司，以及那些需要承担责任的政府机构报销他们服务的资格"（p.442）。

著名的心理治疗执业者道 · 彼得森（Don Peterson）也同意柯瓦斯阿默的警告。他直率地反对许多执业者所持的观点，认为"实征证据总是负面的，但我的经验告诉我，不是这样的，因此我拒绝改变我的疗法"（Peterson，1995，p.977），他说这些人"不会从我这里得到任何同情"（p.977）。简而言之，彼得森认为心理治疗执业者必须对科学证据做出回应，否则这整个行业就会被公众看作是一个"不负责任的"行业（p.977），从而丧失了社会的支持。

在本书的早期版本中，并没有特别强调，心理学家内部的不专业行为和反科学态度，在很大的程度上导致了这个学科的公众形象问题。因此有些读者意见很大，认为我是太容易地放过了心理学家们自己的过失。在这一版中，我特别加强

了对我们这个学科的自我批评，以求平衡我在这一方面的论述。好在，道斯在1994年出版了《纸牌搭成的屋子：建构在虚幻上的心理学和心理治疗》一书，它对我完成这一平衡任务的帮助很大。如果你不相信心理学家自己是造成这个领域困境的重要原因，那你就需要好好地读一下这本书。在这本需要很大勇气才能写成的书里，道斯毫无保留地自曝家丑，并主张在专事研究人类问题的心理学里，采取科学态度对整个社会有很大的实用价值（虽然它的这一潜力至今几乎尚未开始开发）。例如，道斯写道：“现在确实有一门真正的心理科学，这门科学是在许多人多年来为之做了大量工作的基础上建立起来的，但是，这门科学目前正因为一些执业者的不专业行为，而逐渐受到忽视、贬低、和反对。这些执业者只是在嘴皮上认同这门学科而已”（p. ⅶ）。

道斯（还有许多其他人，见Dineen，1996；Lilienfeld，1998；Lilienfeld等，2000；Mook，2001；Watters & Ofshe，1999）所反对的是，这个领域越来越关注所谓的“同业公会”的问题，而非科学研究的问题。行业公会比较关心的是像发放营业执照之类的问题。我们可以从这一学科最古老的专业组织——美国心理学会的各种活动中看出这一端倪。从理论上说，实施营业执照制，为的是要保护通过特殊正规心理学训练所培养出来的专才。但是这个领域现在似乎忘记了它的特殊性在哪里。请你回想一下，我们在本书一开头所作的讨论，心理学领域的独特性在于，它将科学研究的工具用于研究人类的行为上。也就是说，心理学研究并不认可那些运用直觉去“解读”人的手段。确实，有大量的研究文献在支

持这一点。请再回想一下，上一章关于精算预测和临床预测的讨论。对这个问题，40多年来的研究一致表明，在预测几乎任何与心理学相关的事件结果时，相关变量的量化分析总比那些经验丰富的临床医师预测得更准（Dawes et al., 1989; Faust et al., 1988; Goldberg, 1959, 1968, 1991; Meehl, 1954; Sawyer, 1966）。而且，即使是在研究心理治疗的真正（不是预测）疗效这一领域，证据也表明，不管是那些经过培训取得营业执照的心理学家，还是那些拥有多年丰富经验的临床心理学家都不能取得比较好的疗效。许多人对这个问题做了研究之后发现，非心理学专业的执业者（如社会工作专业人士）所取得的疗效与那些取得心理学营业执照的临床心理学家是一样的（Landman & Dawes, 1982; Smith et al., 1980）。事实上，即使那些没有经过任何资历的助理心理专业人士所取得的疗效，也与那些临床心理学家不分伯仲（Berman & Norton, 1985; Christensen & Jacobson, 1994）。最后，也没有证据表明，心理治疗的临床经验会提升疗效。一个临床心理学家的多年工作经验竟然与疗效没有相关（Christensen & Jacobson, 1994; Landman & Dawes, 1982; Smith et al., 1980）。

如果心理学这一领域真的想要遵循依靠实征证据来引导行动之原则的话，那么它早就应该将这些研究结果公诸于世，藉以帮助美国去寻找更便宜的方法来保证全民的心理健康。正如道斯（1994）所言：

这些结果对心理健康领域的政府公共政策有很大的影响。我们不应该浪费资源与钱财，去支持那些收取了高价却不能提

供更有效服务的人（比起那些根本没有经过职业训练的人而言），他们的判断和预测实际上要比用那些依据轻而易得的变量数据所推出的、最简单的统计结论还要糟糕。我们应当正视，心理治疗的疗效是与职业训练或发放执照无关，这一研究发现。我们也应当正视，在精心地进行过标准化的测验上所测得的过去行为及表现的分数才是对未来行为的最佳预测，而不是墨迹测试或由访谈中取得的印象。结论是，为了减轻病人的心理痛苦与折磨，我们应该更多地依靠那些听起来还算科学的、以社区为本的心理健康项目，或依靠助理专业人员的帮助，因为这些人能够让这些病人少花些冤枉钱（p.5）。

心理学领域不但没有顺应这一建议，相反地，他们还为发放营业执照及只准有营业执照的人执业这一行规辩护，声称这样做是为了维护心理学的科学地位，然后再反过来用执照来掩盖执业者内部的不科学行为。例如，一个受过良好训练的心理学家应当知道，我们有把握对总体的行为做出预测，但是在预测具体个人的行为时，就有很大的不确定性（见第十章和第十一章），因此即使是最有能力的心理学家，也不应该在没有强调这点的情况下，去做任何个人预测。正如道斯（1994）所提到的：

一个心理健康专家自信地预测某一个人将来可能会出现的行为（例如，会有暴力行为）时，已经肯定了他不是一个好的专家，因为有研究证实，不论是一个心理健康专家，或者其他什么人，都不可能有充分的证据来支持他做出那么自信的陈述。（专业人士经常声称他们的专业角色“要求”他们做出这样自

信的判断，而不管他们是否真的有那么大的把握。这个辩解是不对的，他们不是被“要求”这样做，是他们“自愿”这样做的。)(p.vii)。

总之，美国心理学会曾经助长了心理治疗领域的这股不正之风。这股风气让人觉得，心理学家能够通过训练，而得到对个体的行为进行“直觉式”洞察的能力。然而，研究证据并不支持这一观点。当被人质疑执照制度是否只是用来包庇这一个行业的低能者时，该组织就把它的科学资历作为令箭。(一位心理学会主席这样回应社会人士对心理学做出的攻击：“我们是以科学为基础的，这就是我们与社会工作者、咨询师和吉普赛卜卦者不同的地方”；道斯，1994，p.21)。但是，这一领域用来维护其科学地位的方法却正好揭示了他们的一个错误的观念，认为有执照的心理学家是那些具有独特的“临床洞察力”的从业者。道斯的书中揭露了部分美国心理学会成员玩弄狡猾的两面派手法。这在一定程度上，导致了1980年代美国心理协会(APS)的成立。这一学会的成员是由那些对美国心理学会那种只关注蓝十字保险公司是否可以报销心理服务费，而不重视科学的做法感到厌倦的心理学家所组成。

史特·利连恩费德(Scott Lilienfield)是一位因为其事业早期对临床心理学所做的贡献，而获得大卫·沙科夫(David Shakow)奖的学者。他曾在颁奖典礼上不断重申以上所述的论点，并警告：“在临床心理学这一领域，我们似乎对处理伪科学这一问题完全没有兴趣，这是一个非常令人吃惊的现象，因为这个问题的火苗已经烧到我们的后院了”

（Lilienfeld，1998，p.3）。他还列出了20世纪90年代在临床心理学领域泛滥成灾的几种伪科学，包括：

1. 用于治疗创伤的那些未经检验的奇特疗法；
2. 已经被证实是无效的、对自闭症的一些疗法，例如辅助沟通术（见第六章）；
3. 继续使用一些未被充分检验的心理评估工具，例如各种投射测验；
4. 各种利用潜意识来做自我治疗的录音带；
5. 使用高度暗示性的治疗技巧来诱导出病人儿时受虐的记忆。

利连恩费德曾引用了著名临床研究者保罗·米赫（Paul Meehl）的话，“如果我们不对这一行业进行清洗整顿，不作我们学生执着科学思考的榜样，外行人很快就会取代了我们”（Meehl，1993，p.728；也见Mahrer，2000）。在同一篇文章中，米赫警告我们，如果临床心理学不采取一个完全科学的立场，那么这些从业者也只不过是收入丰厚的算命先生而已。米赫提到了这样一种趋势（在十一章中已经讨论过），临床执业者往往让人感觉，他们对人类的行为有“特殊”的知识。然而，实征证据却无法支持这一点。临床心理学家应当更注重知识的实征检验，米赫（1993）进一步警告说，“那些自以为拿了博士学位，就能使自己免于抽样、感知、记录、记忆、回忆和推论上错误的人，是非常荒唐和自大的”（p.728）。

心理学家约翰·帕瑞斯（John Perez）认为，道斯、利

连恩费德和米赫所提出的观点可以整合为一个结论，那就是，临床心理学家的自救运动，应当转向谋取病人的支持，而不是那些想怎么治疗就怎么治疗的执业者。正如帕瑞斯（Perez，1999）所言，“我们必须决定，我们是否想制造一个执业者想干嘛就干嘛的宽松环境，甚至牺牲科学证据把关；还是选择去保护病人的权益，以给他们最有效的治疗”（p.205～206）。

然而，心理学领域依然被那些挂出招牌自称为心理学家，却不能坚持遵循这个学科的纪律、暗地里在进行伪科学勾当的人所践踏。当一个领域无法自律并不能排除不科学的做法时，它就不再值得同情了。例如，多伦多的国家邮报曾报道加利福尼亚州的一位心理治疗医师佩里（Perry）用占星术表格来治疗抑郁症的故事，这种占星术显然是十分荒唐的（人们可以通过电话向其咨询，而咨询费为150元/小时）。故事的作者告诉我们，“心理学界可能很想清除像佩里之类的人，但是可能不太容易。佩里有临床心理学的博士学位，他的许多占星术心理学的训练都是基于荣格的研究，这个人到今天仍在大学里任教。” 搞了半天，佩里原来是一个有营业执照的心理治疗师，这个执照是由加利福尼亚州州政府所成立的一个行为科学审查委员会所颁发的，他本人还是加州婚姻与家庭治疗师协会的成员（Milstone，2000）。

不过，事情看似有慢慢好转的迹象。在2002年一本新的杂志诞生了，它就是“对心理健康实践的科学评述”（Lilienfeld，2002）。这本杂志致力于将那些科学的治疗方法与那些伪科学的治疗区别开来，它已经得到科学心理治疗实践

委员会的认可。更令人振奋的是，至少有一些心理学组织已经痛下决心来整顿临床实践，以及消除在实践过程中，所存有的那种"做什么都行"的态度。利连恩费德和洛哈（Lilienfield & Lohr，2000）报告了亚利桑那州心理学资格审查委员会吊销一位心理学家之执照的事件。这个心理学家试图采用伪科学的治疗方法来治疗恐惧症，这种方法是按照预定的顺序轻拍患者身体的各个部位。不必说，这种方法未经控制研究来证实其有效性，亚利桑那州这一委员会命令这位治疗师停止使用这种方法，并且给他"带职查看"的处罚——一个心理学组织为整顿其内部，而严格禁止其成员使用伪科学方法，这一事例在心理学界还是非常罕见的。

简言之，心理学具有像吉柯（Jekyll）和海德（Hyde）那样的阴阳双重人格。极端严密的科学实验与伪科学和反科学的态度并列而存。这一双重人格在20世纪90年代早期关于"记忆恢复—记忆错误"的争论中表现得淋漓尽致（Garry et al.，1999；Loftus，1997；Loftus & Guyer，2002；Loftus & Ketcham，1994；Pezdek & Banks，1996；Pezdek & Hodge，1999；Shermer，1997）。许多病人报告说，他们记起了几十年前，当他们还是小孩的时候，曾经受虐待的片段。这些片段过去一直被遗忘了，但是在治疗过程中，通过干预能够重新被唤醒出来。很明显地，这些记忆中，有一部分是由治疗本身所引发的（Campbell，1998；Loftus & Guyer，2002；Piper，1998）。有些人认为这种记忆是绝对不可信，另外有些人则认为它是可信的。在这个由富有争议的社会虐童问题所制造出的，充满情绪性的氛围下，心理学

家们提供了一些比较理性、平衡的观点，其中比较重要的是，它提供了一些客观的科学证据（Bremner，Shobe & Kihlstrom，2000；Clancy，Schacter，McNally，& Pitman，2000；Pezdek & Banks，1996）。在这里，我们充分地看到心理学的双重人格。在经由治疗干预所引发的错误记忆，甚至“记忆恢复”这一现象的案例中，有一些是由于某些不合格的、对科学无知的治疗师所造成的，而这些治疗师都是临床心理学家（Pezdek & Banks，1996）。从另一个角度来看，尽管我们目前对这场争论的结论还不是很全面、很确定，但是很多有关这方面的研究都是另一些心理学家通过艰辛的努力得出来的成果，而这些心理学家是用实征的方法来研究与“记忆恢复”相关的现象（Pezdek & Banks，1996）。从这里，我们可以看到，心理学制造了一个问题，但同时也在解决这个问题！真是中了“解铃还需系铃人”这句话。

我希望在这一小节的阐述，能够帮我洗清“为心理学家脱罪”的恶名。在本章的开头，我引用了丹杰费尔德的口头禅来做例子，可能令人感觉，我是在维护心理学家，把心理学的形象问题都归罪于他人。心理学家道格拉斯·默克（Douglas Mook）在他的一本关于研究方法的书中提到了，我用丹杰费尔德的玩笑，并且评论道:“确实，通常心理学得不到应有的尊敬，但是，有时为了某些错误的原因，它所得到的尊敬，又完全不是“当之无愧”的”（Mook，2001，p. 473）。我完全同意这一感受。默克是对的，心理学的学生应当知道这个学科所面临的矛盾困境。就像本书中所表述的那样，作为一门研究人类行为的科学，心理学通常没有得到太

多的尊敬。但是，心理学通常传递给公众的形象却是：一位临床心理学家宣称其有独特的、洞察人的能力，而这种洞察能力是不能建立在研究证据的基础上。这时，心理学却又得到了它不应该得的、过多的尊敬。这个学科的知识，通常是由那些不尊重心理学的独特之处的人来传递给公众的。这一独特的特征就是，采取科学的方法来验证关于人类行为的各种解释及说法。

每个人不都是一个心理学家吗？行为的内隐理论

我们每个人都有关于人类行为的一些想法或理论。如果没有这些想法或理论，很难想像我们怎么能活下去。从这一意义上讲，我们都是心理学家，不过是一个个体心理学家。区分出由这种个体心理学和由科学心理学研究所总结出来的知识是十分重要的。这种区别之所以重要，是因为在许多大众读物里，这两者的区别经常是故意地被弄模糊了。让我们看看这是怎么一回事。

我们的个体心理学知识与那些对行为进行科学研究所获得的知识相比，有哪些区别呢？前面已经讨论过一些。我们的个体心理学知识多数是“食谱知识”。我们做某件事，是因为我们认为它会或想要它导致其他人做出某一些相应的行为，或是因为我们相信这些事会帮助我们实现某些目标。这些都是所谓的食谱知识。但是，个体心理学和科学心理学（也包括一些食谱知识）的区分并不在于有或没有食谱知识而已。心

理治疗，在不同程度上，是在提供关于个人和环境的食谱知识，人们可以用之来解决个人的问题或让生活过得更美好一些。其实，这两者最主要的区别是，科学心理学总是力图通过实征检验来验证食谱知识的有效性。

科学评估具有系统性和可控制性，这些特性是个人评估程序所不可能具备的，因此是有缺陷的。事实上，心理学对于决策选择的研究表明，当人类行为发生的情境与他们原有的信念背道而驰时，就很难在其中觉察到任何其他相关关系（见Baron，2000），我们只看到我们想看到的东西。虽然心理学家已经找到出现这种现象的许多原因（Kunda，1999；Lassiter，Geers，Munhall，Ploutz-Snyder，& Breitenbecher，2002；Stanovich，1999，出版中），但是它们并不是本文现在关注的重点，故不在此多谈。如果我们只看到我们想看到的东西，那么，即使我们想在个别人的基础上评估个体食谱知识，那些先入为主的偏见，也会使我们变成一个不称职的行为观察者，从而使我们的评估工作更加困难。科学方法的发明正是要避开这些单一个别的观察者的偏见。结果是，通过科学心理学产生的食谱知识可能会更精确，因为比起个体食谱知识，它们经过更加严格的检验程序。

就像本章前面所讨论的那样，个体心理学和科学心理学之间的差别不仅仅只是在于对食谱知识的检验。科学总是把目标定在一个更高的点上，要在自然世界中找到比食谱知识更多的东西。科学家们想要在食谱知识下面寻求那些更为通用、更为基本的原理。但许多其他人则对人类行为的探究仅限于食谱知识。事实上，对人类行为的根本原因进行深入研究这一想法，

已经让一些人感到害怕了，他们想法总是避开这种知识。

当然，也有些人的个体心理学与科学心理学有相似之处，也积极寻找更基本的心理学原理和理论。但是，这些人的个体理论在许多重要方面与科学理论还是不相同。我们曾经提到过，这是因为这些个体理论是无法证伪的。许多人的个体心理学知识并没有周密及合理的结构，常常只是由一堆适用于当时情境的流行语及时髦话题所堆砌而成，有时这些话还会自相矛盾。它们向人们保证，怎么样都是可以找到一个合理的解释的，而且那些能够推翻这一解释，从而会彻底动摇人们信念的那些反证都是不可能发生的。久之，这堆陈腔滥调就会被无限延伸到可以解释任何发生的事情上了。正如第二章内容中所讨论的，尽管这些理论很具有抚慰人的功能，但是这一类的理论就只能带来心慰，再没有别的了。由于它们只是在事情发生后对之所提出的解释，因此对将来要发生的事情没有任何的预测能力。由于它们不能预测未来，也就没有告诉我们什么有价值的东西。心理学科的理论必须符合可证伪的标准，这就是心理科学与许多外行人的个体心理学的不同之处。心理学理论能够被证明是错误的，这样知识才能不断增长和更新。

有一些个体理论，不仅只是一堆不可以证伪的陈腔滥调，他们确实隐藏着一套心理学理论，而且是以一个比较周密而合理的结构组织起来的，同时它们也是有可能被证伪的。然而，像这样一类的个体心理学理论仍然存在着与个体食谱知识一样的一些困难。虽然它们组织周密合理并且有被证伪的可能性，但是它们仍然会遇到评估实征数据的困难，而这些评估有可能会导致对理论的进一步修改，是科学研究中重要的一

步。一个个体的观察，并不是采用科学方法的操纵及控制方式来建构的。它们会受制于个人带有的偏见，并且是以一种断断续续的方式取得的，因此可能导致在评估时犯概率推理方面的错误（见第十章）。在第六章中看到，尽管人们对移动物体有一年复一年的个人体验，他们有关自然世界中运动的直觉理论，通常都是大错特错的。在变化性更大的人类行为领域，个体理论是错了的可能性就会更大了。

抵制科学心理学的来源

如前所述，千万不要把个体心理学理论和科学心理学的知识混为一谈。这种混淆有时是蓄意制造出来的，目的是要暗中削弱心理学在公众心目中的地位。如果“每个人都是心理学家”是指每一个人都有自己的内隐理论的话，那么这句话是不错的。但是它常常被隐晦地歪曲成，心理学不是一门科学。

第一章中已讨论过，为什么建立一个科学心理学的想法会威胁到某些人。行为科学知识的不断累积，将会改变那些提供心理信息数据来源的个体、群体和组织的本质，从而发现新的心理及行为的规律。很自然地，对那些一直以来以解释及评论人类心理和行为为生的人来说，这些新研究发现必将会令他们丧失他们的权威，因此他们势必要抵制科学心理学，以确保其权威不会受到威胁。在本书的第一章中曾提到过，科学的进步会不断地剥夺那些权威人士的地位。行星的运行、物质的本质、疾病的来源过去曾经是神学家、哲学家和通才作家的领域。现在，天文学、物理学、医学、遗传学和其他学科逐

渐把这些领域给拿走，放在科学家的研究范围之内。

许多宗教都已经逐渐发展到不再敢声称，它们对宇宙结构有专门的知识。除了一些局部性的争议之外，例如对创世说的争论，科学与宗教之间产生巨型大战已经成为历史，因为他们已经各有各职了。科学家们发现自然世界的结构，而许多宗教则对在应用这些发现时可能带来的影响提出评论。宗教已经再也没有与科学家们一争短长的权利了，因为他们已经没有能力再对科学的发现到底是代表什么意义这一点指手划脚了。对自然世界的裁定权，无疑已经掌握在科学家手中了。

之所以会如此，是因为信念评估的标准已经在改变了。很少有新闻报纸会刊登有关土星带构成的、非常有立场的社论文章。为什么呢？并没有什么审查在阻止这类社论的发表呀！很明显的，原因是写这类社论根本没有用。因为公众相信，在这一方面的知识有发言权的是科学家，而不是社论撰写家。仅只是100年前，报纸与那些教坛上的牧师曾对动物世界的物种起源学说大肆抨击。现在，这类评论在很大程度上都消失了。科学已经把任何理性思考者会去相信这些人之评论的客观条件给彻底地摧毁了。心理学则是给自然世界中的另一个巨大领域（人类行为），带来破坏这些条件的威胁。100年后，标题为“早期儿童成长，是遗传还是环境的影响？”的新闻报道可能听起来会很愚蠢、过时，就像我们现在听到大主教雅舍尔声称世界是在公元前4004年形成的时的那种感觉。

有些人发现在面对心理学时，很难接受有一天它会到达像前面所述天文学（土星带结构）那样的地步。他们顽固地坚持他们有权利，对心理学的研究结果评头论足，并发表他们

自已对人类行为的个人看法，即便是他们的意见是与事实相冲突的情况下，也不肯罢休。当然，在这里我们不是讨论权利的问题，因为在一个自由的社会里，每个人都有发言权，不管他/她所说的是否正确。重要的是要意识到，许多人想要的不仅仅只是发表有关人类行为之意见的权利，他们真正想要的是，人们相信他们说的话的必要条件。当他们谈论一些人类心理学问题时，他们希望周围的环境有利于人们接受他们的想法。这就是为什么认为心理学是"做什么都行"的说法会有大量支持者的原因，所谓"做什么都行"，就是暗含心理学的说法是不能由实征方法来决定对错的，它只是一个视角不同的问题。科学对于这种"一切看情境而定"的观点是一种威胁，因为它有一系列严格的标准和程序，用以确定哪些说法是可信的。总之，在科学面前，各种观点、想法都是平等的，都必须经过并通过检验。这种将错误理论排除的能力推动了科学不断地前进。

简言之，许多对科学心理学的抵制来自利益上的冲突。在前面几章中已讨论过，许多伪科学已经发展成为数以百万美元计的产业，它们利用了公众没有意识到行为也可以用实征方证来检验的这一事实。（在美国市场上，占星算命先生的数量是天文学家的20倍；见Gilovich，1991，p.2）公众也没有察觉支撑这些产业的那些技术及治疗方法，如星相预测、潜意识减肥、生物规律、苦杏仁苷（laetrile，一种抗肿瘤药）、通灵手术，都已被实征检验过，并被证实是无效的。美国众议会下面的一个委员会曾估计过，人们每年大约花100亿美金在这些骗人的治疗上，这与花在正规的医疗研究上的经费相比那

真是大巫见小巫了（Eisenberg et al., 1993；美国国会，1984）。

许多伪科学和虚假疗法的散播者就是在心理学领域这种"做什么都行"的氛围中滋长起来的。它提供了一个利用公众无知的良好环境，因为如果"做什么都行"，公众的消费者权益就得不到保障了。正如律师彼特·哈勃（Peter Huber）所言，"在科学的边缘和科学之外……形形色色相信顺势疗法药物（homeopathic medicine）、水晶和金字塔之疗效的人……必须诽谤正规的科学，借以为他们的异端邪说提供立足之地"（Huber，1990，p.97）。那些兜售伪科学的人总是为了自身的利益，刻意地去掩盖，事实上确实是有一套科学机制可以用于检验行为理论，这一事实。吉瑟林（1989）警告说，"这里的道理其实很简单，人们试图推销他们的观点，而那些真正能评估那些观点之好坏的人，不是那些购买他们的产品的人"（p.139）。在行为理论和治疗这一领域，心理学家就是那些知道如何来评估产品的人。这就是为什么伪心理学产业极力反对科学心理学有对行为理论进行评估的权威地位。然而，伪科学的散播者通常不需要与心理学家展开正面的斗争。他们采用迂回进攻的战略，他们直接走向媒体，直接向公众推销他们的理论或疗法。大众传媒给那些想要绕过科学心理学的狂徒、骗子和伪科学提供了一个方便。一大堆的电视脱口秀节目在20世纪90年代应运而生，在这些节目里，嘉宾并没有被要求提供他们言论之科学证据及其参考文献。只要这些嘉宾能"吸引人看"，他们就可以上电视了。

民间常识通常包括许多妄想：人们力图相信世界正在朝着

他们所期望的方向发展，而实际情况可能并非如此。为此，科学家们经常必须做那些吃力不讨好的工作，去告诉公众，这个世界并不是像他们所想像的那样（"快餐对你的健康没有好处"）。媒体本来可以帮助科学家来完成这一艰难的工作（告诉人们真相，而不是迎合他们的期望），然而，它们却把重心放在"娱乐"而不是"事实"的供给上，从而把事情搞得更糟糕（Postman，1985，1999；Postman & Powers，1992）。

科学确实把那些不符合最低检验标准的、自称为是特别知识的理论及疗法清除出局。在本书里，曾经很简短的探讨过，在科学领域中，充分检验是什么意思，不充分的检验又是什么意思。内省、个人经验和见证叙述都被认为是不充分的检验。那些运用此类证据，以支持其观点的人，经常是非心理学界的评论家，就会对此产生不满，因为他们早在当代心理学还未诞生前，就已经在使用这些不充分的证据了，并没有什么不妥。

请不要以为我想让科学心理学承担一个阴沉严肃、让人扫兴的角色。恰恰相反，科学心理学的研究发现其实要比起那些整天在媒体上叽叽喳喳、反反复复的伪科学言论要有趣和精彩得多。同时，也不应该认为科学家是反对幻想、想像及欲念的——但它们必须在适当的情境下出现。米德瓦（1990）认为：

我坚信人们在适当的情境下可以随心所欲地夸夸其谈。我相信多数人在心理上需要像保罗·詹宁斯（Paul Jennings）所说的那样有"Bunkrapt"（在床上冥想或幻想）。（你也许还记得詹宁斯在打字时，连"Bankrupt""破产"都打成了

“Bunkrapt”。）任何人都需要幻想，我喜欢在听维格纳的歌剧或者读托柯金的小说时幻想。但是，可不能将这个习惯带到科学里面（p.5）。

如果我们停下来，思考一分钟，大多数人都会赞同米德瓦的观点。我们进入歌剧院或电影院时，经常会陷入奇思妙想，但这种情形却不太可能发生在我们去看病、买保险、到托儿所给孩子们注册、坐飞机，或者修理汽车的时候。这种情形也不太可能发生在我们进行心理治疗，或请教育心理学家去测试自己学习不佳的孩子，或是把自己的朋友带到大学心理诊所进行自杀干预的时候。心理学在追求真相的过程中，必须像其他学科一样，把那些毫无依据的胡思乱想、“常识”、商业广告卖点、宗教意见、见证和妄想都清除出去。

一门科学必须告诉人们，他们的想法和意见是有用的，但不是在这里！这是一件相当困难的事。心理学最近正是面临这样一个困境。这与心理学产生的历史时期是有关的。大多数学科成熟于精英控制社会结构的年代，在那个时代，普通人的意见没有影响力。而心理学则产生于一个民主的传媒时代，忽视公众意见十分危险。许多心理学家正在努力改善心理学与公众在沟通方面非常糟糕的记录。当越来越多的心理学家参与到这项工作中时，势必会增加与那帮将个人心理学和科学心理学混为一谈的人冲突。

虽然每个人都有一套直觉的物理理论，但不是每个人都是物理学家。如果我们放弃让这些霸占了科学物理学的个人物理理论，就是为物理世界的真正科学研究做了一件好事，而这

门学科的成果最终反过来会为我们所有人做好事，因为科学是属于大众的。同样地，不是人人都是心理学家，但心理科学所发现的事实和理论随时都可以拿来应用，并丰富我们对所有人的理解。

结　束　语

现在已经到了整本书的尾声。本书对如何正视心理学只提供了一个素描，但它可以对你理解心理学这门学科是如何运作的，以及评估新的心理学理论，有很大的帮助。我们的素描勾勒出以下几个线条：

1. 心理学的进展是通过研究可解决的实征问题来取得的。心理学是由许多不同的子领域构成，某些领域的问题要比其他领域更困难，因此心理学的发展是不平衡的。
2. 心理学家提出可证伪的理论来解释他们的研究发现。
3. 理论中的概念是用可操作的程序来定义的，这些定义将随着证据的积累而逐渐变化。
4. 这些理论是通过系统实征的方法来检验的，用这种方法收集来的数据是公开的，也就是说，它允许其他科学家重复这些实验并提出批评。
5. 心理学家的数据和理论，只有在那些经过同行评审程序的科学杂志上发表之后，才算是进入科学领域。
6. 实征主义之所以具有系统性，是因为它符合真正实验研究的两个特点——遵循控制和操纵的逻辑。
7. 心理学家采用许多不同的方法来取得他们的结论，这些

方法的优点和缺点各有不同。

8. 在许多情况下，只有经过许多实验资料的缓慢积累才能得出结论。

9. 最终发现的行为规律可以说总是一种概率关系。

1961年，英国心理学家博德本特说过一段话，这段话到现在仍然与在当时一样的切题。对我们来说，它可以当作是教我们如何正视心理学的一个精辟总结：

我们在此带着一份质疑告终，我们不确定将来的心理学家会持有些什么信念。事情本来就该如此。没有人可以坐在高背椅里就能够抓住事实的本质。在做新的实验之前，我们理所当然不知道结果会是怎么样的。不需要那些经常在讲坛、新闻报道和学校的颁奖典礼上听到的、关于人性的教条来引导我们。正相反，我们必须做好准备，去容忍目前尚不完善的、有关行为的知识，但却还要坚信客观方法的威力，总有一天，会让我们得到那些知识的（p.200～201）。

（林升栋　译）

参 考 文 献

Adler, J. E. (1998, January). Open minds and the argument from ignorance. *Skeptical Inquirer, 22*(1), 41–44.

Alcock, J. E. (1981). *Parapsychology: Science or magic?* Oxford, England: Pergamon Press.

Alcock, J. E. (1984). Parapsychology's past eight years: A lack-of-progress report. *Skeptical Inquirer, 8*, 312–320.

Alcock, J. E. (1987). Parapsychology: Science of the anomalous or search for the soul? *Behavioral and Brain Sciences, 10*, 553–643.

Alcock, J. E. (1990). *Science and supernature: A critical appraisal of parapsychology.* Buffalo, NY: Prometheus Books.

Alloy, L. B., Abramson, L. Y., & Francis, E. L. (1999). Do negative cognitive styles confer vulnerability to depression? *Current Directions in Psychological Science, 8*, 128–132.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Anderson, C. A., & Anderson, K. B. (1996). Violent crime rate studies in philosophical context: A destructive testing approach to heat and Southern culture of violence effects. *Journal of Personality and Social Psychology, 70*, 740–756.

Anderson, C. A., Benjamin Jr., A. J., & Bartholow, B. D. (1998). Does the gun pull the trigger? Automatic priming effects of weapon pictures and weapon names. *Psychological Science, 9*, 308–314.

Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology, 78*, 772–790.

Anderson, C. A., & Gunderson, K. (1991). *Paul E. Meehl: Selected philosophical and methodological papers.* Minneapolis: University of Minnesota Press.

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science, 8*, 3–9.

Anderson, D. R., Huston, A. C., Schmitt, K. L., Linebarger, D. L., & Wright, J. C. (2001). Early childhood television viewing and adolescent behavior. *Monographs of the Society for Research in Child Development, 66*(1) (Serial No. 264), 1–158.

Angell, M., & Kassirer, J. P. (1998). Alternative medicine: The risks of untested and unregulated remedies. *The New England Journal of Medicine, 339*(12), 839–841.

Arkes, H. R., & Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin, 125*, 591–600.

Asimov, I. (1989). The relativity of wrong. *Skeptical Inquirer, 14*, 35–44.

Auletta, K. (1992). *Three blind mice: How the TV networks lost their way.* New York: Vintage Books.

Austin, E. J., & Deary, I. J. (2002). Personality dispositions. In R. J. Sternberg (Ed.), *Why smart people can be so stupid* (pp. 187–211). New Haven, CT: Yale University Press.

Azar, B. (1998, March). Are psychologists shooting themselves in the foot? *APA Monitor*, pp. 18–19.

Azar, B. (1999, November). Crowder mixes theories with humility. *APA Monitor*, p. 18.

Bachman, J. G., & Schulenberg, J. (1993). How part-time work intensity relates to drug use, problem behavior, time use, and satisfaction among high school seniors: Are these consequences or merely correlates? *Developmental Psychology, 29*, 220–235.

Bainbridge, W. S., & Stark, R. (1980). Superstitions: Old and new. *Skeptical Inquirer, 4*, 18–21.

Banaji, M. R., & Crowder, R. G. (1989). The bankruptcy of everyday memory. *American Psychologist, 44*, 1185–1193.

Baron, J. (1998). *Judgment misguided: Intuition and error in public decision making.* New York: Oxford University Press.

Baron, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge, MA: Cambridge University Press.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: The MIT Press.

Barrett, L., Dunbar, R., & Lycett, J. (2002). *Human evolutionary psychology*. Princeton, NJ: Princeton University Press.

Baumeister, R. F. (1999, January). Low self-esteem does not cause aggression. *APA Monitor*, p. 7.

Baumeister, R. F., Boden, J. M., & Smart, L. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review, 103*, 5–33.

Baumeister, R. F., Bushman, B. J., & Campbell, K. W. (2000). Self-esteem, narcissism, and aggression: Does violence result from low self-esteem or from threatened egotism? *Current Directions in Psychological Science, 9*, 26–29.

Bazerman, M. (1999). *Smart money decisions*. New York: John Wiley.

Bazerman, M. (2001). Consumer research for consumers. *Journal of Consumer Research, 27*, 499–504.

Beatty, J. (1996, May). What election '96 should be about. *The Atlantic Monthly*, pp. 114–120.

Beck, A. R., & Pirovano, C. M. (1996). Facilitated communications' performance on a task of receptive language with children and youth with autism. *Journal of Autism and Developmental Disorders, 26*, 497–512.

Belsky, G., & Gilovich, T. (1999). *Why smart people make big money mistakes—And how to correct them: Lessons from the new science of behavioral economics*. New York: Simon & Schuster.

Ben-Shakhar, G., Bar-Hillel, M., Blui, Y., Ben-Abba, E., & Flug, A. (1989). Can graphological analysis predict occupational success? *Journal of Applied Psychology, 71*, 645–653.

Berkowitz, L., & Donnerstein, E. (1982). External validity is more than skin deep. *American Psychologist, 37*, 245–257.

Berliner, D. C., & Biddle, B. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Reading, MA: Addison Wesley.

Berman, J., & Norton, N. (1985). Does professional training make a therapist more effective? *Psychological Bulletin, 98*, 401–407.

Berra, T. M. (1990). *Evolution and the myth of creation*. Stanford, CA: Stanford University Press.

Bersoff, D. (1981). Testing and the law. *American Psychologist, 36*, 1047–1056.

Bersoff, D. (1982). Larry P. and PASE: Judicial report cards on the validity of intelligence tests. In T. R. Kratochwill (Ed.), *Advances in school psychology* (pp. 145–162). Hillsdale, NJ: Erlbaum.

Best, J. (1982). Misconceptions about psychology among students who perform highly. *Psychological Reports, 51*, 239–244.

Beyerstein, B. L. (1999). Whence cometh the myth that we use only 10% of our brains? In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 3–24). Chichester, England: John Wiley & Sons.

Birnbaum, M. H. (1999). Testing critical properties of decision making on the Internet. *Psychological Science, 10*, 399–407.

Bjorklund, D. F. (1995). *Children's thinking: Developmental function and individual differences* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Bjorklund, D. F., & Pellegrini, A. D. (2000). Child development and evolutionary psychology. *Child Development, 71*, 1687–1708.

Bjorklund, D. F., & Pellegrini, A. D. (2002). *The origins of human nature: Evolutionary developmental psychology*. Washington DC: APA.

Boden, M. A. (1990). *The creative mind: Myths and mechanisms*. New York: Basic Books.

Bok, S. (1974). The ethics of giving placebos. *Scientific American, 23*, 17–23.

Boneau, C. A. (1990). Psychological literacy: A first approximation. *American Psychologist, 45*, 891–900.

Bornstein, B. H., & Rajki, M. (1994). Extra-legal factors and product liability: The influence of mock jurors' demographic characteristics and intuitions about the cause of an injury. *Behavioral Sciences and the Law, 12*, 127–147.

Bornstein, B. H., & Zickafoose, D. J. (1999). "I know I know it, I know I saw it": The stability of the confidence-accuracy rela-

tionship across domains. *Journal of Experimental Psychology: Applied, 5,* 76–88.

Bornstein, R. F. (1989). Exposure and affect. *Psychological Bulletin, 106,* 265–289.

Bower, B. (1990, July 21). The ticcing link. *Science News, 138,* 42–44.

Bower, B. (1996a, August 31). Brain images illuminate Tourette syndrome. *Science News, 150,* 133.

Bower, B. (1996b, August 24). New pitch for placebo power. *Science News, 150,* 123.

Boyer, P. J. (1999, May 17). Big guns. *The New Yorker,* pp. 54–67.

Braun, P., & Yaniv, I. (1992). A case study of expert judgment: Economists' probabilities versus base-rate model forecasts. *Journal of Behavioral Decision Making, 5,* 217–231.

Breitmeyer, B. J., & Ramey, C. T. (1986). Biological nonoptimality and quality of postnatal environment as codeterminants of intellectual development. *Child Development, 57,* 1151–1165.

Bremner, J. D., Shobe, K. K., & Kihlstrom, J. F. (2000). False memories in women with self-reported childhood sexual abuse: An empirical study. *Psychological Science, 11,* 333–337.

Brenneman, R. (1990). *Deadly blessings: Faith healing on trial.* Buffalo, NY: Prometheus Books.

Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in probability and frequency judgments: A critical examination. *Organizational Behavior and Human Decision Processes, 65,* 212–219.

Broadbent, D. E. (1958). *Perception and communication.* New York: Pergamon Press.

Broadbent, D. (1961). *Behaviour.* New York: Basic Books.

Broadbent, D. (1973). *In defense of empirical psychology.* London: Methuen.

Bronfenbrenner, U., & Mahoney, M. (1975). The structure and verification of hypotheses. In U. Bronfenbrenner & M. Mahoney (Eds.), *Influence on human development* (pp. 131–152). Hinsdale, IL: Dryden.

Bronfenbrenner, U., McClelland, P., Wethington, E., Moen, P., & Ceci, S. J. (1996). *The state of Americans.* New York: Free Press.

Broniarczyk, S., & Alba, J. W. (1994). Theory versus data in prediction and correlation tasks. *Organizational Behavior and Human Decision Processes, 57,* 117–139.

Bronowski, J. (1956). *Science and human values.* New York: Harper & Row.

Bronowski, J. (1973). *The ascent of man.* Boston: Little, Brown.

Bronowski, J. (1974). Science, poetry, and human specificity. *American Scholar, 43,* 386–404.

Bronowski, J. (1977). *A sense of the future.* Cambridge, MA: MIT Press.

Bronowski, J. (1978a). *The common sense of science.* Cambridge, MA: Harvard University Press.

Bronowski, J. (1978b). *Magic, science, and civilization.* New York: Columbia University Press.

Budiansky, S. (1984, April 22). The meat of the matter. *Washington Post Book World,* p. 7.

Bunge, M. (1983). Speculation: Wild and sound. *New Ideas in Psychology, 1,* 3–6.

Burgess, C. A., Kirsch, I., Shane, H., Niederauer, K., Graham, S., & Bacon, A. (1998). Facilitated communication as an ideomotor response. *Psychological Science, 9,* 71–74.

Bushman, B. J., & Anderson, C. A. (2002). Violent video games and hostile expectations: A test of the general aggression model. *Personality and Social Psychology Bulletin, 28*(1679–1686).

Buss, D. M. (1985). Human mate selection. *American Scientist, 73,* 47–51.

Buss, D. M. (1999). *Evolutionary psychology: The new science of the mind.* Boston: Allyn and Bacon.

Buss, D. M. (2000). The evolution of happiness. *American Psychologist, 55,* 15–23.

Buss, D. M. (2001). Cognitive biases and emotional wisdom in the evolution of conflict between the sexes. Current Directions in *Psychological Science, 10,* 219–223.

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of Personality and Social Psychology, 50,* 559–570.

Byrnes, G., & Kelly, I. (1992). Crisis calls and lunar cycles: A twenty-year review. *Psychological Reports, 71,* 779–795.

Byrnes, J. P. (2001). *Minds, brains, and learning: Understanding the psychological and educational relevance of neuroscientific research.* New York: Guilford Press.

Campbell, T. (1998). *Smoke and mirrors: The devastating effect of false sexual abuse claims*. New York: Plenum.

Cardon, L. R., Smith, S., Fulker, D., Kimberling, W., Pennington, B., & DeFries, J. (1994). Quantitative trait locus for reading disability on chromosome 6. *Science, 266*, 276–279.

Carpenter, S. (1999, August 21). Kansas cuts evolution from curriculum. *Science News, 156*, 117.

Carroll, M., & Nelson, T. O. (1993). Failure to obtain a generation effect during naturalistic learning. *Memory and Cognition, 21*, 361–366.

Casscells, W., Schoenberger, A., & Graboys, T. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine, 299*, 999–1001.

Cassidy, J. (1995, October 15). Who killed the middle class? *New Yorker*, pp. 113–124.

Catrambone, R., Jones, C. M., Jonides, J., & Seifert, C. (1995). Reasoning about curvilinear motion: Using principles or analogy. *Memory and Cognition, 23*, 368–373.

Ceci, S. J., & Bruck, M. (1993a). Child witnesses: Translating research into policy. *Social Policy Report: Society for Research in Child Development, 7*(3), 1–30.

Ceci, S. J., & Bruck, M. (1993b). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113*, 403–439.

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom*. Washington, DC: American Psychological Association.

Ceci, S. J., & Hembrooke, H. (Eds.). (1998). *Expert witnesses in child abuse cases*. Washington, DC: American Psychological Association.

Chang, E. C. (1996). Cultural differences in optimism, pessimism, and coping: Predictors of subsequent adjustment in Asian American and Caucasian American college students. *Journal of Counseling Psychology, 43*, 113–123.

Chase, A. (1977). *The legacy of Malthus*. New York: Knopf.

Cheng, C. (2001). Assessing coping flexibility in real-life and laboratory settings: A multimethod approach. *Journal of Personality and Social Psychology, 80*, 814–833.

Choi, I., Nisbett, R. E., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin, 125*, 47–63.

Christensen, A., & Jacobson, N. S. (1994). Who (or what) can do psychotherapy: The status and challenge of nonprofessional therapies. *Psychological Science, 5*, 8–14.

Christensen, D. (1999, July 17). Does practice make perfect? The benefits of busy hospitals. *Science News, 156*, 44–45.

Christensen, D. (2001, February 3). Medicinal mimicry: Sometimes, placebos work—but how? *Science News*, pp. 74–78.

Churchland, P. M. (1988). *Matter and consciousness* (rev. ed.). Cambridge, MA: MIT Press.

Churchland, P. M. (1995). *The engine of reason, the seat of the soul*. Cambridge, MA: MIT Press.

Cialdini, R. B. (1997). Professionally responsible communication with the public: Giving psychology away. *Personality and Social Psychology Bulletin, 23*, 675–683.

Clancy, S. A., Schacter, D. L., McNally, R. J., & Pitman, R. K. (2000). False recognition in women reporting recovered memories of sexual abuse. *Psychological Science, 11*, 26–31.

Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science*. New York: Oxford University Press.

Clotfelter, C., & Cook, P. (1989). *Selling hope: State lotteries in America*. Cambridge, MA: Harvard University Press.

Clotfelter, C., & Cook, P. (1993). The "gambler's fallacy" in lottery play. *Management Science, 39*, 1521–1525.

Coan, J. A. (1997). Lost in a shopping mall: An experience with controversial research. *Ethics & Behavior, 7*, 271–284.

Coates, W., Jehle, D., & Cottington, E. (1989). Trauma and the full moon: A waning theory. *Annals of Emergency Medicine, 18*, 763–765.

Cohen, R. (1985, December 17). Despair: America's double standard. *Detroit Free Press*, p. 11.

Cole, K. C. (1998). Calculated risks. *Skeptical Inquirer, 22*(5), 32–36.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.),

The adapted mind (pp. 163–228). New York: Oxford University Press.

Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed.) (pp. 91–115). New York: Guilford Press.

Cournaud, A. (1977). The code of the scientist and its relationship to ethics. *Science, 198,* 699–705.

Cowley, G., & Biddle, N. A. (1994, June 20). Waving away the pain. *Newsweek,* pp. 70–71.

Crease, R. P., & Samios, N. P. (1991, January). Managing the unmanageable. *The Atlantic Monthly,* pp. 80–88.

Crews, F. (1993, November 18). The unknown Freud. *New York Review of Books,* pp. 55–66.

Crews, F. (1996). The verdict on Freud. *Psychological Science, 7,* 63–68.

Crews, F. (Ed.). (1998). *Unauthorized Freud: Doubters confront a legend.* New York: Viking.

Culver, R., Rotton, J., & Kelly, I. W. (1988). Moon mechanisms and myths. *Psychological Reports, 62,* 683–710.

Cummins, R. A., & Prior, M. P. (1992). Autism and assisted communication: A response to Bilken. *Harvard Educational Review, 62,* 228–241.

Curtis, B. M., & O'Keefe, J. H. (2002). Autonomic tone as a cardiovascular risk factor: The dangers of chronic fight or flight. *Mayo Clinic Proceedings, 77*(45–54).

Davidow, J., & Levinson, E. M. (1993). Heuristic principles and cognitive bias in decision making: Implications for assessment in school psychology. *Psychology in the Schools, 30,* 351–361.

Davis, D., & Holt, C. (1993). *Experimental economics.* Princeton, NJ: Princeton University Press.

Dawda, D., & Martin, J. (2001). Psychologists on psychology: The inquiry beliefs of psychologists by subdiscipline, gender, and age. *Review of General Psychology, 5,* 163–179.

Dawes, R. M. (1988). *Rational choice in an uncertain world.* San Diego: Harcourt Brace & Jovanovich.

Dawes, R. M. (1991). Probabilistic versus causal thinking. In D. Cicchetti & W. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* (Vol. 1, pp. 235–264). Minneapolis: University of Minnesota Press.

Dawes, R. M. (1994). *House of cards: Psychology and psychotherapy built on myth.* New York: Free Press.

Dawes, R. M. (2001). *Everyday irrationality.* Boulder, CO: Westview Press.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668–1673.

Dawkins, R. (1998). *Unweaving the rainbow.* Boston: Houghton Mifflin.

Day, S. X., & Rounds, J. (1998). Universality of vocational interest structure among racial minorities. *American Psychologist, 53,* 728–736.

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain.* Oxford: Oxford University Press.

Deary, I. J. (2001). Human intelligence differences: Towards a combined experimental-differential approach. *Trends in Cognitive Sciences, 5*(4), 164–170.

Della Sala, S. (Ed.). (1999). *Mind myths: Exploring popular assumptions about the mind and brain.* Chichester, England: John Wiley & Sons.

Dembroski, T. M., & Costa, P. T. (1987). Coronary prone behavior: Components of the Type A pattern and hostility. *Journal of Personality, 55,* 211–235.

Dembroski, T., & Costa, P. (1988). Assessment of coronary-prone behavior: A review. *Annals of Behavioral Medicine, 10,* 60–63.

Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life.* New York: Simon & Schuster.

Dennett, D. C. (1998). *Brainchildren: Essays on designing minds.* Cambridge, MA: MIT Press.

Dennett, D. C. (2000). Why getting it right matters: Postmodernism and truth. *Free Inquiry, 20*(1), 40–43.

Dickson, D. H., & Kelly, I. W. (1985). The "Barnum effect" in personality assessment: A review of the literature. *Psychological Reports, 57,* 367–382.

Dillon, K. (1993, Spring). Facilitated communication, autism, and ouija. *Skeptical Inquirer, 17,* 281–287.

Dineen, T. (1996). *Manufacturing victims: What the psychology industry is doing to people.* Montreal: Robert Davies.

diSessa, A. A. (1996). What do "just plain folk" know about physics? In D. R. Olson & N. Torrance (Eds.), *Handbook of education and human development: New models of learning, teaching, and schooling* (pp. 709–730). Cambridge, MA: Blackwell Publishers.

Dixon, R., & Backman, L. (Eds.). (1995). *Compensating for psychological deficits and declines: Managing losses and promoting gain.* Mahwah, NJ: Erlbaum.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *American Biology Teacher, 35,* 125–129.

Dodes, J. E. (1997). The mysterious placebo. *Skeptical Inquirer, 21*(1), 44–45.

Doherty, M. E., Chadwick, R., Garavan, H., Barr, D., & Mynatt, C. R. (1996). On people's understanding of the diagnostic implications of probabilistic data. *Memory and Cognition, 24,* 644–654.

Domjan, M., & Purdy, J. E. (1995). Animal research in psychology: More than meets the eye of the general psychology student. *American Psychologist, 50,* 496–503.

Dos Passos, J. (1964). *Occasions and protests.* New York: Henry Regnery.

Dowie, J., & Elstein, A. (Eds.). (1988). *Professional judgment: A reader in clinical decision making.* New York: Cambridge University Press.

Druckman, D., & Swets, J. A. (1988). *Enhancing human performance: Issues, theories and techniques.* Washington, DC: National Academy Press.

Dudley, U. (1998). Numerology: Comes the revolution. *Skeptical Inquirer, 22*(5), 29–31.

Ehrenberg, R. G., Brewer, D. J., Gamoran, A., & Williams, J. D. (2001). Class size and student achievement. *Psychological Science in the Public Interest, 2,* 1–27.

Ehri, L. C., Nunes, S., Stahl, S., & Willows, D. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's Meta-Analysis. *Review of Educational Research, 71,* 393–447.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50,* 387–395.

Eisenberg, L. (1977). The social imperatives of medical research. *Science, 198,* 1105–1110.

Eisenberg, D. M., Kessler, R., Foster, C., Norlock, F., Calkins, D., & Delbanco, T. (1993). Unconventional medicine in the United States. *The New England Journal of Medicine, 328*(4), 246–252.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Ernst, E., & Abbot, N. C. (1999). I shall please: The mysterious power of placebos. In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 209–213). Chichester, England: John Wiley & Sons.

Estes, W. (1979). Experimental psychology: An overview. In E. Hearst (Eds.), *The first century of experimental psychology* (pp. 12–38). Hillsdale, NJ: Erlbaum.

Fackelmann, K. (1996, November 9). Gastrointestinal blues. *Science News, 150,* 302–303.

Fagley, N. (1988). Judgmental heuristics: Implications for the decision making of school psychologists. *School Psychology Review, 17,* 311–321.

Falk, R. (1989). Judgment of coincidences: Mine versus yours. *American Journal of Psychology, 102,* 477–493.

Falk, R., & MacGregor, D. (1983). The surprisingness of coincidences. In P. Humphreys, O. Svenson, & A. Vari (Eds.), *Analysis and aiding decision processes* (pp. 489–502). Budapest: Akademiai Kiado.

Fancher, R. (1985). *The intelligence men.* New York: W. W. Norton.

Fantino, E., & Esfandiari, A. (2002). Probability matching: Encouraging optimal responding in humans. *Canadian Journal of Experimental Psychology, 56,* 58–63.

Farber, I. B., & Churchland, P. S. (1995). Consciousness and the neurosciences: Philosophical and theoretical issues. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1295–1306). Cambridge, MA: MIT Press.

Faust, D., Hart, T., Guilmette, T., & Arkes, H. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Profes-*

sional Psychology: Research and Practice, 19, 578–582.

Feeney, D. (1987). Human rights and animal welfare. *American Psychologist, 42,* 593–599.

Fernald, D. (1984). *The Hans legacy.* Hillsdale, NJ: Erlbaum.

Ferrari, M., & Sternberg, R. J. (1998). The development of mental abilities and styles. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology* (Vol. 2, pp. 899–946). New York: John Wiley.

Finkel, A. M. (1996, May). Who's exaggerating? *Discover, 17*(5), 48–54.

Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21,* 97–109.

Foster, E. A., Jobling, M. A., Taylor, P. G., Donnelly, P., Deknijff, P., Renemieremet, J., Zerjal, T., & Tyler-Smith, C. (1998). Jefferson fathered slave's last child. *Nature, 396,* 27–28.

Foster, K. R., & Huber, P. W. (1999). *Judging science: Scientific knowledge and the federal courts.* Cambridge, MA: MIT Press.

Fox, R. E. (1996). Charlatanism, scientism, and psychology's social contract. *American Psychologist, 51,* 777–784.

Frank, R. H. (1999). *Luxury fever: Why money fails to satisfy in an era of excess.* New York: Free Press.

Frank, R. H., & Cook, P. J. (1995). *The winner-take-all society.* New York: Free Press.

Fridson, M. S. (1993). *Investment illusions.* New York: John Wiley.

Fried, S. B. (1994). *American popular psychology: An interdisciplinary research guide.* New York: Garland Publishing.

Fried, S. B. (1998). An undergraduate course in American popular psychology. *Teaching of Psychology, 25,* 38–40.

Fried, S. B., & Schultis, G. A. (1995). *The best self-help and self awareness books.* Chicago: American Library Association.

Friedman, M., & Booth-Kewley, S. (1987). The "disease-prone personality." *American Psychologist, 41,* 539–555.

Friedman, M., & Ulmer, D. (1984). *Treating Type A behavior and your heart.* New York: Knopf.

Friedrich, J., Buday, E., & Kerr, D. (2000). Statistical training in psychology: A national survey and commentary on undergraduate programs. *Teaching of Psychology, 27,* 248–257.

Friedrich-Cofer, L., & Huston, A. (1986). Television violence and aggression: The debate continues. *Psychological Bulletin, 100,* 364–371.

Frith, U. (1989). *Autism: Explaining the enigma.* Oxford, England: Basil Blackwell.

Frith, U. (1993, June). Autism. *Scientific American, 268*(6), 108–114.

Gage, N. L. (1996). Confronting counsels of despair for the behavioral sciences. *Educational Researcher, 25*(3), 5–15.

Gage, N. L., & Berliner, D. (1984). *Educational psychology* (3rd ed.). San Francisco: Freeman.

Gal, I., & Baron, J. (1996). Understanding repeated simple choices. *Thinking and Reasoning, 2,* 81–98.

Galaburda, A. (1994). Developmental dyslexia and animal studies: At the interface between cognition and neurology. *Cognition, 50,* 133–149.

Garb, H. N., Florio, C. M., & Grove, W. M. (1998). The validity of the Rorschach and the Minnesota Multiphasic Personality Inventory. *Psychological Science, 9,* 402–404.

Garb, H. N., & Schramke, C. J. (1996). Judgment research and neuropsychological assessment: A narrative review and meta-analysis. *Psychological Bulletin, 120,* 140–153.

Gardner, M. (1972, April). Mathematical games: Why the long arm of coincidence is usually not as long as it seems. *Scientific American, 227*(4), 110–112.

Gardner, M. (2001). Facilitated communication: A cruel farce. *Skeptical Inquirer, 25*(1), 17–19.

Garry, M., Frame, S., & Loftus, E. F. (1999). Lie down and tell me about your childhood. In S. Della Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 113–124). Chichester, England: John Wiley & Sons.

Gawande, A. (1998, February 8). No mistake. *The New Yorker,* pp. 74–81.

Gawande, A. (1999, February 9). The cancer-cluster myth. *The New Yorker,* pp. 35–37.

Gazzaniga, M. S. (2000). *The new cognitive neurosciences* (2nd ed.). Cambridge, MA: MIT Press.

Geschwind, N. (1985). The biology of dyslexia. In D. Gray & J. Kavanagh (Eds.), *Behavioral measures of dyslexia* (pp. 19–37). Parkton, MD: York Press.

Ghiselin, M. T. (1989). *Intellectual compromise*. New York: Paragon House.

Gibson, E. J. (1994). Has psychology a future? *Psychological Science, 5*, 69–76.

Gigerenzer, G. (1984). External validity of laboratory experiments: The frequency-validity relationship. *American Journal of Psychology, 97*, 185–195.

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York: Free Press.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*, 295–314.

Gladwell, M. (1996, July 8). Conquering the coma. *The New Yorker*, pp. 34–40.

Glassner, B. (1999). *The culture of fear: Why Americans are afraid of the wrong things*. New York: Basic Books.

Gleitman, H. (1981). *Psychology*. New York: W. W. Norton.

Goldberg, L. R. (1959). The effectiveness of clinicians' judgments: The diagnosis of organic brain damage from the Bender Gestalt Test. *Journal of Consulting Psychology, 23*, 25–33.

Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgments. *American Psychologist, 23*, 483–496.

Goldberg, L. R. (1991). Human mind versus regression equation: Five contrasts. In D. Cicchetti & W. Grove (Eds.), *Thinking clearly about psychology: Essays in honor of Paul E. Meehl* (Vol. 1, pp. 173–184). Minneapolis: University of Minnesota Press.

Goldstein, M., & Goldstein, I. (1978). *How we know: An exploration of the scientific process*. New York: Plenum Press.

Goldstein, W. M., & Hogarth, R. M. (Eds.). (1997). *Research on judgment and decision making*. Cambridge, England: Cambridge University Press.

Goswami, U. (1998). *Cognition in children*. Hove, England: Psychology Press.

Gould, S. J. (1987). Justice Scalia's misunderstanding. *Natural History, 96*, 14–21.

Greenberger, D., & Steinberg, L. (1986). *When teenagers work*. New York: Basic Books.

Greenwald, A. G., Spangenberg, E., Pratkanis, A., & Eskenazi, J. (1991). Double-blind tests of subliminal self-help audiotapes. *Psychological Science, 2*, 119–122.

Grice, G. (2001, July 30). Slice of life: How a convicted killer's corpse brought anatomy into the digital age. *The New Yorker*, pp. 36–41.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411–435.

Griggs, R. A., Proctor, D. L., & Bujak-Johnson, A. (2002). The nonexistent common core. *American Psychologist, 57*, 452–453.

Grimmer, M. R. (1992). Searching for security in the mystical: The function of paranormal beliefs. *Skeptical Inquirer, 16*, 173–176.

Groopman, J. (1999, May 10). Pet scan. *The New Yorker*, pp. 46–50.

Hacking, I. (1975). *The emergence of probability*. Cambridge, England: Cambridge University Press.

Hacohen, M. C. (2000). *Karl Popper: The formative years, 1902–1945*. Cambridge, England: Cambridge University Press.

Hagen, M. A. (2001). Damaged goods? What, if anything, does science tell us about the long-term effects of childhood sexual abuse? *Skeptical Inquirer, 25*(1), 54–59.

Haggbloom, S. J., Warnick, R., Warnick, J. E., Jones, V. K., Yarbrough, G. L., Russell, T. M., Borecky, C. M., McGahhey, R., Powell, J. L., Beavers, J., & Monte, E. (2002). The 100 most eminent psychologists of the 20th century. *Review of General Psychology, 6*, 139–152.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, training, and metacognitive monitoring. *American Psychologist, 53*, 449–455.

Hammond, K. R. (1996). *Human judgement and social policy*. New York: Oxford University Press.

Hammond, K. R., Harvey, L. O., & Hastie, R. (1992). Making better use of scientific knowledge: Separating truth from justice. *Psychological Science, 3*, 80–87.

Harlow, H. F. (1958). The nature of love. *American Psychologist, 13*, 673–685.

Harlow, H. F., & Suomi, S. J. (1970). The nature of love—Simplified. *American Psychologist, 25*, 161–168.

Harrington, A. (1997). *The placebo effect*. Cambridge, MA: Harvard University Press.

Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world*. Thousand Oaks, CA: Sage.

Hearst, E. (1979). *The first century of experimental psychology*. Hillsdale, NJ: Erlbaum.

Heath, L., Tindale, R., Edwards, J., Posavac, E., Bryant, F., Henderson-King, E., Suarez-Balcazar, Y., & Myers, J. (Eds.). (1994). *Applications of heuristics and biases to social issues*. New York: Plenum Press.

Heinsman, D. T., & Shadish, W. (1996). Assignment methods in experimentation: When do nonrandomized experiments approximate answers from randomized experiments? *Psychological Methods, 1*, 154–169.

Higbee, K. L., & Clay, S. L. (1998). College students' beliefs in the ten-percent myth. *The Journal of Psychology, 132*, 469–476.

Hines, T. M. (1988). *Pseudoscience and the paranormal*. Buffalo, NY: Prometheus Books.

Hines, T. M. (1998). Comprehensive review of biorhythm theory. *Psychological Reports, 83*, 19–64.

Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science in the Public Interest, 3*, 39–77.

Holton, G. (1996). *Einstein, history, and other passions: The rebellion against science at the end of the twentieth century*. Reading, MA: Addison-Wesley.

Holton, G., & Roller, D. (1958). *Foundations of modern physical science*. Reading, MA: Addison Wesley.

Hooft, G. T. (2000). Physics and the paranormal. *Skeptical Inquirer, 24*(2), 27–30.

Huber, P. W. (1990). *Pathological science in court*. Daedalus, 119, 97–118.

Hudson, A., Melita, B., & Arnold, N. (1993). A case study assessing the validity of facilitated communication. *Journal of Autism and Developmental Disorders, 23*, 165–173.

Huesmann, L. R., & Eron, L. D. (1986). *Television and the aggressive child: A cross-national comparison*. Hillsdale, NJ: Erlbaum.

Humphrey, N. (1996). *Leaps of faith*. New York: Copernicus.

Hunt, M. (1999). *The new know-nothings: The political foes of the scientific study of human nature*. New Brunswick, NJ: Transaction Publishers.

Hyman, R. (1981). Cold reading: How to convince strangers that you know all about them. In K. Frazier (Eds.), *Paranormal borderlands of science* (pp. 79–96). Buffalo, NY: Prometheus Books.

Hyman, R. (1992). What does Goodfellow's classic analysis tell researchers today? *Journal of Experimental Psychology: General, 121*, 128–129.

Hyman, R. (1996). Evaluation of the military's twenty-year program on psychic spying. *Skeptical Inquirer, 20*(2), 21–26.

Hynd, G. W., Clinton, A., & Hiemenz, J. R. (1999). The neuropsychological basis of learning disabilities. In R. J. Sternberg & L. Spear-Swerling (Eds.), *Perspectives on learning disabilities* (pp. 60–79). New York: Westview/HarperCollins.

If it sounds just ducky, it probably quacks. (1990, March 23). *Detroit Free Press*.

Immen, W. (1996, August 8). Could you repeat that in Klingon? *Globe & Mail (Toronto)*.

Jacobson, J. W., Mulick, J. A., & Schwartz, A. A. (1995). A history of facilitated communication: Science, pseudoscience, and antiscience. *American Psychologist, 50*, 750–765.

Jacobson, J. W., Mulick, J. A., & Schwartz, A. A. (1996). If a tree falls in the woods. *American Psychologist, 51*, 988–989.

Johnson, H. (1991). *Sleepwalking through history: America in the Reagan years*. New York: Anchor Books.

Kahne, J. (1996). The politics of self-esteem. *American Educational Research Journal, 33*, 3–22.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T.

Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (Eds.). (2000). *Choices, values, and frames*. Cambridge, England: Cambridge University Press.

Kalat, J. W. (1995). *Biological psychology* (5th ed.). Pacific Grove, CA: Brooks/Cole.

Kaminer, W. (1999). *Sleeping with extra-terrestrials: The rise of irrationalism and the perils of piety*. New York: Pantheon.

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research. *American Psychologist, 56*, 405–416.

Kelly, I. W. (1997). Modern astrology: A critique. *Psychological Reports, 81*, 931–962.

Kelly, I. W. (1998). Why astrology doesn't work. *Psychological Reports, 82*, 527–546.

Kelly, I., Rotton, J., & Culver, R. (1985). The moon was full and nothing happened: A review of studies on the moon and human behavior and lunar beliefs. *Skeptical Inquirer, 10*, 129–143.

Kenrick, D. T. (2001). Evolutionary psychology, cognitive science, and dynamical systems: Building an integrative paradigm. *Current Directions in Psychological Science, 10*(1), 13–17.

Kershaw, A. (1991, July 21). *The psychic guide to hiring and firing*. London Independent.

Kety, S. (1974). Preface. In J. Swazey, *Chlorpromazine in psychiatry: A study of therapeutic intervention*. Cambridge, MA: MIT Press.

Kim, J. (1994, September). Watch out for investing newsletters luring you with outdated returns. *Money Magazine*, pp. 12–13.

Kimble, G. A. (1999). Functional behaviorism: A plan for unity in psychology. *American Psychologist, 54*, 981–988.

King, F. (1993, March). Science literacy and public support of science. *American Psychological Society Observer, 6*(2), 2–11.

King, R. N., & Koehler, D. J. (2000). Illusory correlations in graphological inference. Journal of Experimental Psychology: *Applied, 6*, 336–348.

Kirsch, I. (Ed.). (1999). *How expectancies shape experience*. Washington, DC: American Psychological Association.

Knight, G. P., Johnson, L., Carlo, G., & Eisenberg, N. (1994). A multiplicative model of the dispositional antecedents of a prosocial behavior: Predicting more of the people more of the time. *Journal of Personality and Social Psychology, 66*, 178–183.

Koehler, J. J. (1993). The normative status of base rates at trial. In N. J. Castellan (Eds.), *Individual and group decision making* (pp. 137–149). Hillsdale, NJ: Erlbaum.

Kohn, A. (1990). *You know what they say . . . : The truth about popular beliefs*. New York: HarperCollins.

Kolb, B., & Whishaw, I. (1990). *Fundamentals of human neuropsychology* (3rd ed.). New York: Freeman.

Kopta, S. M., Lueger, R. J., Saunders, S. M., & Howard, K. I. (1999). Individual psychotherapy outcome and process research. *Annual Review of Psychology, 50*, 441–469.

Kramer, S. P. (1987). *How to think like a scientist: Answering questions by the scientific method*. New York: Thomas Y. Crowell Junior Books.

Krauthammer, C. (1985, December 29). The twilight of psychotherapy? *Detroit News*.

Krippner, S. (1977). *Advances in parapsychological research: 1. Psychokinesis*. New York: Plenum Press.

Krull, D. S., Hui-Min Loy, M., Lin, J., Wang, C., Chen, S., & Zhao, X. (1999). The fundamental attribution error: Correspondence bias in individualist and collectivist cultures. *Personality and Social Psychology Bulletin, 25*, 1208–1219.

Kuhn, D., Weinstock, M., & Flaton, R. (1994). How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological Science, 5*, 289–296.

Kunda, Z. (1999). *Social cognition: Making sense of people*. Cambridge, MA: MIT Press.

Kushner, H. I. (1999). *A cursing brain? The histories of Tourette Syndrome*. Cambridge, MA: Harvard University Press.

Landman, J. T., & Dawes, R. M. (1982). Psychotherapy Outcome. *American Psychologist, 31*, 504–516.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32*, 311–328.

Langewiesche, W. (1993, December). The turn. *The Atlantic Monthly*, pp. 115–122.

Lassiter, G. D., Geers, A. L., Munhall, P. J., Ploutz-Snyder, R. J., & Breitenbecher, D. L. (2002). Illusory causation: Why it occurs. *Psychological Science, 13*, 299–305.

Latane, B., & Darley, J. (1970). *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts.

Lazar, I., Darlington, R., Murray, H., Royce, J., & Sniper, A. (1982). Lasting effects of early education: A report from the Consortium of Longitudinal Studies. *Monographs of the Society for Research in Child Development, 47* (Serial No. 195).

Leary, D. E. (2001). One big idea, one ultimate concern. *American Psychologist, 56*, 425–432.

Leavy, J. (1992). Our spooky presidential coincidences contest. *Skeptical Inquirer, 16*, 316–319.

Lee, V., Brooks-Gunn, J., Schnur, E., & Liaw, F. (1990). Are Head Start's effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Child Development, 61*, 495–507.

Lees-Haley, P. R. (1997). Manipulation of perception in mass tort litigation. *Natural Resources & Environment, 12*, 64–68.

Leibowitz, H. W. (1996). The symbiosis between basic and applied research. *American Psychologist, 51*, 366–370.

Leli, D., & Filskov, S. (1981). Clinical-actuarial detection and description of brain impairment with the W-B Form 1. *Journal of Clinical Psychology, 37*, 623–629.

Lerner, M. J., & Miller, D. T. (1978). Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin, 85*, 1030–1051.

Leslie, A. M. (2000). "Theory of mind" as a mechanism of selective attention. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.) (pp. 1235–1247). Cambridge, MA: MIT Press.

Levin, J. R., & O'Donnell, A. M. (2000). What to do about educational research's credibility gaps? *Issues in Education: Contributions from Educational Psychology, 5*, 1–87.

Levine, K., Shane, H., & Wharton, R. (1994). What if . . . : A plea to professionals to consider the risk-benefit ratio of facilitated communication. *Mental Retardation, 32*, 300–304.

Lewis, M. (1997). *Trail fever*. New York: Knopf.

Li, C. (1975). *Path analysis: A primer*. Pacific Grove, CA: Boxwood Press.

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. Journal of Experimental Psychology: *Human Learning and Memory, 4*, 551–578.

Lilienfeld, S. O. (1996). EMDR treatment: Less than meets the eye? *Skeptical Inquirer, 20*(1), 25–31.

Lilienfeld, S. O. (1998). Pseudoscience in contemporary clinical psychology: What it is and what we can do about it. *The Clinical Psychologist, 51*(4), 3–9.

Lilienfeld, S. O. (1999). Projective measures of personality and psychopathology: How well do they work? *Skeptical Inquirer, 23*(5), 32–39.

Lilienfeld, S. O. (2002). New journal dedicated to distinguishing science from pseudoscience in mental health. *Skeptical Inquirer, 26*(3), 7–8.

Lilienfeld, S. O., & Lohr, J. M. (2000). Thought field therapy practitioners and educators sanctioned. *Skeptical Inquirer, 24*(2), 5.

Lilienfeld, S. O., Lohr, J. M., & Moirer, D. (2001). The teaching of courses in the science and pseudoscience of psychology: Useful resources. *Teaching of Psychology, 28*(3), 182–191.

Lilienfeld, S. O.,Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1*, 27–66.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181–1209.

Loftus, E. F. (1993). Psychologists in the eyewitness world. *American Psychologist, 48*, 550–552.

Loftus, E. F. (1994). The repressed memory controversy. *American Psychologist, 49*, 443–445.

Loftus, E. F. (1997, September). Creating false memories. *Scientific American, 276*(9), 70–75.

Loftus, E. F., & Guyer, M. J. (2002, May/June). Who abused Jane Doe: The hazards of the single case history. *Skeptical Inquirer, 26*(3), 24–32.

Loftus, E. F., & Ketcham, K. (1994). *The myth of repressed memory: False memories and allegations of sexual abuse.* New York: St. Martin's Press.

Lopes, L., & Oden, G. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 392–400.

MacCoun, R. (2002, December). Why a psychologist won the Nobel Prize in economics. *APS Observer, 15*(10), 1–8.

MacDonald, A. (1990, October 3). Parents fear wrong things, survey suggests. *Ann Arbor News* (taken from The New York Times wire service).

MacDougall, C. D. (1983). *Superstition and the press.* Buffalo, NY: Prometheus Books.

Macmillan, M. (1997). *Freud evaluated.* Cambridge, MA: MIT Press.

Magee, B. (1985). *Philosophy and the real world: An introduction to Karl Popper.* LaSalle, IL: Open Court.

Mahrer, A. R. (2000). Philosophy of science and the foundations of psychotherapy. *American Psychologist, 55,* 1117–1125.

Malkiel, B. G. (1999). *A random walk down Wall Street.* New York: W. W. Norton.

Margolis, H. (1996). *Dealing with risk.* Chicago: University of Chicago Press.

Marinoff, L. (1999). *Plato, not Prozac: Applying philosophy to everyday problems.* New York: HarperCollins.

Market Facts, I. (1984). *1983 consumer research study on reading and book purchasing.* New York: Book Industry Study Group.

Marks, D. F. (2001). *The psychology of the psychic.* Buffalo, NY: Prometheus Books.

Marks, D., & Kammann, R. (1980). *The psychology of the psychic.* Buffalo, NY: Prometheus Books.

Marshall, E. (1980). Police science and psychics. *Science, 210,* 994–995.

Martin, B. (1998). Coincidences: Remarkable or random? *Skeptical Inquirer, 22*(5), 23–28.

Maynard Smith, J. (1998). *Evolutionary genetics* (2nd ed.). Oxford: Oxford University Press.

McBride-Chang, C., & Kail, R. V. (2002). Cross-cultural similarities in the predictors of reading acquisition. *Child Development, 73,* 1392–1407.

McCloskey, M. (1983, April). Intuitive physics. *Scientific American, 248*(4), 122–130.

McCorduck, P. (1979). *Machines who think.* San Francisco: Freeman.

McCullough, M. L. (2001). Freud's seduction theory and its rehabilitation: A saga of one mistake after another. *Review of General Psychology, 5,* 3–22.

McCutcheon, L. E., Furnham, A., & Davis, G. (1993). A cross-national comparison of students' misconceptions about psychology. *Psychological Reports, 72,* 243–247.

McFall, R. M., & Treat, T. A. (1999). Quantifying the information value of clinical assessments with signal detection theory. *Annual Review of Psychology, 50,* 215–241.

McGraw, K. O.,Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science, 11,* 502–506.

Medawar, P. B. (1967). *The art of the soluble.* London: Methuen.

Medawar, P. B. (1979). *Advice to a young scientist.* New York: Harper & Row.

Medawar, P. B. (1984). *The limits of science.* New York: Harper & Row.

Medawar, P. B. (1990). *The threat and the glory.* New York: HarperCollins.

Medawar, P. B., & Medawar, J. S. (1983). *Aristotle to zoos: A philosophical dictionary of biology.* Cambridge, MA: Harvard University Press.

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and review of the literature.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment, 50,* 370–375.

Meehl, P. E. (1991). Law and the fireside inductions: Some reflections of a clinical psychologist. In C. Anderson & K. Gunderson (Eds.), *Paul E. Meehl: Selected philosophical and methodological papers* (pp. 440–480). Minneapolis: University of Minnesota Press.

Meehl, P. E. (1993). Philosophy of science: Help or hindrance? *Psychological Reports, 72,* 707–733.

Menon, T., Morris, M. W., Chiu, C., & Hong, Y. (1999). Culture and the construal of agency: Attribution to individual versus group dispositions. *Journal of Personality and Social Psychology, 76,* 701–717.

Metalsky, G. I., & Joiner, T. E. (1992). Vulnerability to depressive symptomatology: A prospective test of the diathesis-stress and causal mediation components of the hopelessness theory of depression. *Journal of Personality and Social Psychology, 63,* 667–675.

Miller, M., Azrael, D., & Hemmingway, D. (2002). Firearm availability and unintentional firearm deaths, suicide, and homicide among 5–14 year olds. *Journal of Trauma, Injury, Infection, and Critical Care, 52,* 267–275.

Miller, N. E. (1985a). Rx: Biofeedback. *Psychology Today, 19,* 54–59.

Miller, N. E. (1985b). The value of behavioral research on animals. *American Psychologist, 40,* 423–440.

Miller, T. Q., Turner, C., Tindale, R., Posavac, E., & Dugoni, B. (1991). Reasons for the trend toward null findings in research on Type A behavior. *Psychological Bulletin, 110,* 469–485.

Milstone, C. (2000, February 28). California psychotherapist uses astrological charts to treat stress and depression. *National Post (Toronto),* p. D3.

Milton, J., & Wiseman, R. (1999). Does psi exist? Lack of replication of an anomalous process of information transfer. *Psychological Bulletin, 125,* 387–391.

Mishel, L. (1995, Fall). Rising tides, sinking wages. *American Prospect,* pp. 60–64.

Monk, R. (1990). *Ludwig Wittgenstein: The duty of genius.* New York: Free Press.

Mook, D. G. (1982). *Psychological research: Strategies and tactics.* New York: Harper & Row.

Mook, D. G. (1983). In defense of external invalidity. *American Psychologist, 38,* 379–387.

Mook, D. G. (1989). The myth of external validity. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 25–43). Cambridge, England: Cambridge University Press.

Mook, D. G. (2001). *Psychological research: The ideas behind the methods.* New York: Norton.

Moore, T. E. (1995). Subliminal self-help auditory tapes: An empirical test of perceptual consequences. *Canadian Journal of Behavioural Science, 27,* 9–20.

Moore, T. E. (1996). Scientific consensus and expert testimony: Lessons from the Judas Priest trial. *Skeptical Inquirer, 20*(6), 32–38.

Mulick, J., Jacobson, J., & Kobe, F. (1993, Spring). Anguished silence and helping hands: Autism and facilitated communication. *Skeptical Inquirer, 17,* 270–280.

Murchison, C. (1934). *Handbook of general experimental psychology.* Worcester, MA: Clark University Press.

Murray, D., Schwartz, J., & Lichter, S. R. (2001). *It ain't necessarily so: How media make and unmake the scientific picture of reality.* Lanham, MD: Rowman & Littlefield.

Murstein, B. (1980). Mate selection in the 1970s. *Journal of Marriage and Family, 42,* 51–66.

Nathan, P. E., & Gorman, J. M. (Ed.). (1998). *A guide to treatments that work.* New York: Oxford University Press.

National Safety Council. (1990). *Accident facts* (1990 edition). Chicago: Author.

National Safety Council (2001). *Report on Injuries in America, 2001.* Retrieved March 29, 2002, from www.nsc.org/library/rept2000.htm.

National Science Board (2001). Science indicators 2000: Belief in the paranormal or pseudoscience. *Skeptical Inquirer, 25*(1), 12–15.

Neter, E., & Ben-Shakhar, G. (1989). Predictive validity of graphological inferences: A meta-analysis. *Personality and Individual Differences, 10,* 737–745.

Nickerson, R. (1992). *Looking ahead: Human factors challenges in a changing world.* Hillsdale, NJ: Erlbaum.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2,* 175–220.

Nickerson, R. S. (1999). Basic versus applied research. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 409–423). Cambridge, MA: MIT Press.

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review, 109,* 330–357.

Niemeyer, W., & Starlinger, I. (1981). Do the blind hear better? *Audiology, 20,* 503–515.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108,* 291–310.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

Olson, R. K. (1999). Genes, environment, and reading disabilities. In R. J. Sternberg & L. Spear-Swerling (Eds.), *Perspectives on learning disabilities* (pp. 3–21). New York: Westview/HarperCollins.

Olson, R. K., & Forsberg, H. (1993). Disabled and normal readers' eye movements in reading and nonreading tasks. In D. M. Willows, R. Kruk, & E. Corcos (Eds.), *Visual processes in reading and reading disabilities* (pp. 377–391). Hillsdale, NJ: Erlbaum.

Orkin, M. (2000). *What are the odds? Chance in eveyday life.* New York: W. H. Freeman.

Page, E., & Keith, T. (1981). Effects of U.S. private schools: A technical analysis of two recent claims. *Educational Researcher, 10,* 7–17.

Paik, H., & Comstock, G. (1994). The effects of television violence on antisocial behavior: A meta-analysis. *Communications Research, 21,* 516–546.

Paloutzian, R. F. (1983). *Invitation to the psychology of religion.* Glenview, IL: Scott, Foresman.

Paul, A. M. (2001, March/April 2001). Self-help: Shattering the myths. *Psychology Today,* pp. 62–68.

Paul, G. L. (1966). *Insight vs. desensitization in psychotherapy.* Stanford, CA: Stanford University Press.

Paul, G. L. (1967). Insight vs. desensitization in psychotherapy two years after termination. *Journal of Consulting Psychology, 31,* 333–348.

Paulos, J. A. (1988). *Innumeracy: Mathematical illiteracy and its consequences.* New York: Vintage Books.

Pearl, D., Bouthilet, L., & Lazar, J. (Eds.). (1982). *Television and behavior: Ten years of scientific progress and implications for the eighties.* Rockville, MD: National Institute of Mental Health.

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist, 54,* 741–754.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 37,* 51–87.

Pennock, R. (1999). *Tower of Babel: The new creationism.* Cambridge, MA: MIT Press.

Perez, J. E. (1999). Clients deserve empirically supported treatments, not romanticism. *American Psychologist, 54,* 205–206.

Peterson, D. R. (1995). The reflective educator. *American Psychologist, 50,* 975–983.

Pettit, G. S., Bates, J. E., Dodge, K. A., & Meece, D. W. (1999). The impact of after-school peer contact on early adolescent externalizing problems is moderated by parental monitoring, perceived neighborhood safety, and prior adjustment. *Child Development, 70,* 768–778.

Pezdek, K., & Banks, W. P. (Eds.). (1996). *The recovered memory/false memory debate.* San Diego: Academic Press.

Pezdek, K., & Hodge, D. (1999). Planting false childhood memories in children: The role of event plausibility. *Child Development, 70,* 887–895.

Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence.* Cambridge, MA: MIT Press.

Piattelli-Palmarini, M. (1994). *Inevitable illusions: How mistakes of reason rule our minds.* New York: John Wiley.

Pigliucci, M. (2002). Hypothesis testing and the nature of skeptical investigations. *Skeptical Inquirer, 26*(6), 27–48.

Pinker, S. (1997). *How the mind works.* New York: W. W. Norton.

Pinker, S. (2002). *The blank slate: The modern denial of human nature.* New York: Viking.

Piper, A. (1998). Multiple personality disorder: Witchcraft survives in the twentieth century. *Skeptical Inquirer, 22*(3), 44–50.

Poon, L. W., Rubin, D. C., & Wilson, B. A. (Eds.). (1989). *Everyday cognition in adulthood and late life.* Cambridge, England: Cambridge University Press.

Popper, K. R. (1959). *The logic of scientific discovery.* New York: Harper & Row.

Popper, K. R. (1963). *Conjectures and refutations*. New York: Harper & Row.

Popper, K. R. (1972). *Objective knowledge*. Oxford, England: Oxford University Press.

Popper, K. R. (1976). *Unended quest: An intellectual biography*. La Salle, IL: Open Court.

Postman, N. (1988). *Conscientious Objections*. New York: Vintage Books.

Poundstone, W. (1999). *Carl Sagan: A life in the cosmos*. New York: Henry Holt and Co.

Powell, B. (1993, December). Sloppy reasoning, misused data. *Phi Delta Kappan, 75*(4), 283, 352.

Powell, B., & Steelman, L. C. (1996). Bewitched, bothered, and bewildering: The use and misuse of state SAT and ACT scores. *Harvard Educational Review, 66*, 27–59.

Pressley, M. (2002). *Reading instruction that works: The case for balanced teaching* (2nd ed.). New York: Guilford Press.

Prioleau, L., Murdock, M., & Brody, N. (1983). An analysis of psychotherapy versus placebo studies. *Behavioral and Brain Sciences, 6*, 275–310.

Prior, M. P., & Cummins, R. A. (1992). Questions about facilitated communication and autism. *Journal of Autism and Developmental Disorders, 22*, 331–338.

Proctor, R. W., & Capaldi, E. J. (2001). Empirical evaluation and justification of methodologies. *Psychological Bulletin, 127*, 759–772.

Rabbitt, P. (1993). Does it all go together when it goes? *Quarterly Journal of Experimental Psychology, 46A*, 385–434.

Radford, B. (1999). The ten-percent myth. *Skeptical Inquirer, 23*(2), 52–53.

Ramey, S. L. (1999). Head Start and preschool education. *American Psychologist, 54*, 344–346.

Randi, J. (1983). The project alpha experiment: 1. The first two years. *Skeptical Inquirer, 7*, 24–33.

Randi, J. (1986). *Flim-flam*. Buffalo, NY: Prometheus Books.

Randi, J. (1987). *The faith healers*. Buffalo, NY: Prometheus Books.

Randi, J. (1995). *An encyclopedia of claims, frauds, and hoaxes of the occult and supernatural exposed by James Randi*. New York: St. Martin's Press.

Raymo, C. (1999). *Skeptics and true believers*. Toronto: Doubleday Canada.

Rayner, K. (1998). Eye movements in reading and information processing: 20 Years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*, 31–74.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2002, March). How should reading be taught? *Scientific American, 286*(3), 84–91.

Redding, R. E. (1998). *How common-sense psychology can inform law and psycholegal research*. University of Chicago Law School Roundtable, 5, 107–142.

Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *The New England Journal of Medicine, 336*(7), 453–458.

Redelmeier, D. A., & Tibshirani, R. J. (2001). Car phones and car crashes: Some popular misconceptions. *Canadian Medical Association Journal, 164*(11), 1581–1582.

Ridley, M. (1996). *Evolution* (2nd ed.). Cambridge, MA: Blackwell Science.

Ridley, M. (1999). *Genome: The autobiography of a species in 23 chapters*. New York: HarperCollins.

Rind, B., Tromovitch, P., & Bauserman, R. (2001). The validity and appropriateness of methods, analyses, and conclusions in Rind et al. (1998): A rebuttal of victimological critique from Ondersma et al. (2001) and Dallam et al. (2001). *Psychological Bulletin, 127*, 734–758.

Roberts, A. H., Kewman, D. G., Mercier, L., & Hovell, M. (1993). The power of nonspecific effects in healing: Implications for psychological and biological treatments. *Clinical Psychology Review, 13*, 375–391.

Robins, R. W., & Craik, K. H. (1994). A more appropriate test of the Kuhnian displacement thesis. *American Psychologist, 49*, 815–816.

Robins, R. W., Gosling, S. D., & Craik, K. H. (1999). An empirical analysis of trends in psychology. *American Psychologist, 54*, 117–128.

Robins, R. W.,Gosling, S. D., & Craik, K. H. (2000). Trends in psychology: An empirical issue. *American Psychologist, 55*, 276–277.

Robinson, D. N. (2001). Sigmund Koch—Philosophically speaking. *American Psychologist, 56*, 420–424.

Robinson, W. S. (1992). *Computers, minds, and robots.* Philadelphia: Temple University Press.

Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 46*, 775–776.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology.* Philadelphia: Temple University Press.

Rotton, J., & Kelly, I. W. (1985). Much ado about the full moon. *Psychological Bulletin, 97*, 286–306.

Rowe, D. C., Vazsonyi, A. T., & Flannery, D. J. (1994). No more than skin deep: Ethnic and racial similarity in developmental process. *Psychological Review, 101*, 396–413.

Rowe, W. (1993). Psychic detectives: A critical examination. *Skeptical Inquirer, 17*, 159–165.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD Triad Hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology, 76*, 574–586.

Ruggiero, V. (2000). Bad attitude: Confronting the views that hinder students' learning. *American Educator, 24*(2), 10–15.

Ruscio, J. (2002, March). The emptiness of holism. *Skeptical Inquirer, 26*(2), 46–50.

Ruscio, J. (2000). Risky business: Vividness, availability, and the media paradox. *Skeptical Inquirer, 24*(2), 22–26.

Ruse, M. (1999). *Mystery of mysteries: Is evolution a social construction?* Cambridge, MA: Harvard University Press.

Russo, E. (2002, December 9). The biological basis of the placebo effect. *The Scientist, 16*(24), 30–31.

Russo, F. (1999, May). The clinical-trials bottleneck. *The Atlantic Monthly*, pp. 30–36.

Rutter, M. (1979). Maternal deprivation, 1972–1978: New findings, new concepts, and new approaches. *Child Development, 50*, 283–305.

Rzewnicki, R., & Forgays, D. G. (1987). Recidivism and self-cure of smoking and obesity: An attempt to replicate. *American Psychologist, 42*, 97–100.

Sagan, C. (1996). *The demon-haunted world: Science as a candle in the dark.* New York: Random House.

Salthouse, T. A. (1996). Constraints on theories of cognitive aging. *Psychonomic Bulletin & Review, 3*, 287–299.

Salthouse, T. A., & Miles, J. D. (2002). Aging and time-sharing aspects of executive control. *Memory & Cognition, 30*, 572–582.

Sandoval, J., & Miille, M. (1980). Accuracy of judgments of WISC-R item difficulty for minority groups. *Journal of Consulting and Clinical Psychology, 48*, 249–253.

Santrock, J. W., Minnett, A. M., & Campbell, B. D. (1994). *The authoritative guide to self-help books.* New York: Guilford Press.

Sawyer, J. (1996). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66*, 178–200.

Scarborough, D., & Sternberg, S. (Eds.). (1998). *Methods, models, and conceptual issues: An invitation to cognitive science* (Vol. 4). Cambridge, MA: MIT Press.

Schachter, S. (1982). Recidivism and self-cure of smoking and obesity. *American Psychologist, 37*, 436–444.

Schraw, G., & Nietfeld, J. (1998). A further test of the general monitoring skill hypothesis. *Journal of Educational Psychology, 90*, 236–248.

Scott, S. (1999, January 2). Risking all on alternative cancer therapies. *National Post (Toronto)*, p. B1.

Seligmann, J., & Chideya, F. (1992, September 21). Horror story or big hoax? *Newsweek*, p. 75.

Senechal, M., & LeFevre, J. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*, 445–460.

Shaffer, L. (1981). The growth and limits of recipe knowledge. *Journal of Mind and Behavior, 2*, 71–83.

Shane, H. (1993, June). FC: Facilitated or "factitious" communication. *Communicating Together, 11*(2), 11–13.

Shanks, D. R. (1995). Is human learning rational? *Quarterly Journal of Experimental Psychology, 48A*, 257–279.

Shapiro, A. (1960). A contribution to a history of the placebo effect. *Behavioral Science, 5*, 109–135.

Shapiro, A., Shapiro, E., Bruun, R., & Sweet, R. (1978). *Gilles de la Tourette syndrome.* New York: Raven Press.

Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education: Contributions from Educational Psychology, 1*, 1–57.

Shaywitz, S. E. (1996). Dyslexia. *Scientific American, 275*(5), 98–104.

Shepard, R. (1983). "Idealized" figures in textbooks versus psychology as an empirical science. *American Psychologist, 38*, 855.

Shermer, M. (1997). *Why people believe weird things.* New York: W.H. Freeman.

Shontz, F., & Green, P. (1992). Trends in research on the Rorschach: Review and recommendations. *Applied and Preventive Psychology, 1*, 149–156.

Siegert, R. J., & Ward, T. (2002). Clinical psychology and evolutionary psychology: Toward a dialogue. *Review of General Psychology, 6*, 235–259.

Sigman, M., & Ruskin, E. (1999). Continuity and change in the social competence of children with autism, Down syndrome, and developmental delays. *Monographs of the Society for Research in Child Development, 64*(Serial No. 256), 1–114.

Simmons, R., Burgeson, R., Carlton-Ford, S., & Blyth, D. (1987). The impact of cumulative change in early adolescence. *Child Development, 58*, 1220–1234.

Simpson, R. L., & Myles, B. S. (1995). Effectivenesss of facilitated communication with children and youth with autism. *Journal of Special Education, 28*, 424–439.

Sivak, M., & Flannagan, M. J. (2003). Flying and driving after the September 11 attacks. *American Scientist, 91*, 6–7.

Slemrod, J., & Bakija, J. (1996). *Taxing ourselves: A citizen's guide to the great debate over tax reform.* Cambridge, MA: The MIT Press.

Smith, M., & Belcher, R. (1993). Facilitated communication with adults with autism. *Journal of Autism and Developmental Disorders, 23*, 175–183.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy.* Baltimore: Johns Hopkins University Press.

Smith, R. E., Wheeler, G., & Diener, E. (1975). Faith without works. *Journal of Applied Social Psychology, 5*, 320–330.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children.* Washington, DC: National Academy Press.

Solso, R., & Massaro, D. W. (1996). *The science of the mind.* New York: Oxford University Press.

Spanos, N. P. (1996). *Multiple identities and false memories: A sociocognitive perspective.* Washington, DC: American Psychological Association.

Spitz, H. H. (1997). *Nonconscious movements: From mystical messages to facilitated communication.* Mahwah, NJ: Erlbaum.

Stankov, L., & Spilsbury, G. (1978). The measurement of auditory abilities of blind, partially sighted, and sighted children. *Applied Psychological Measurement, 2*, 491–503.

Stanovich, K. E. (1989). Implicit philosophies of mind: The dualism scale and its relationships with religiosity and belief in extrasensory perception. *Journal of Psychology, 123*, 5–23.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning.* Mahwah, NJ: Erlbaum.

Stanovich, K. E. (2000). *Progress in understanding reading.* New York: Guilford Press.

Stanovich, K. E. (in press). *Rationality and evolution: The robot's rebellion in the age of Darwin.* Chicago: University of Chicago Press.

Stanovich, K. E., & West, R. F. (1998). Who uses base rates and P(D/~H)? An analysis of individual differences. *Memory & Cognition, 28*, 161–179.

Steen, L. A. (1990). Numeracy. *Daedalus, 119*, 211–231.

Steinberg, L., Brown, B. B., & Dornbusch, S. M. (1996). *Beyond the classroom: Why school reform has failed and what parents need to do.* New York: Simon & Schuster.

Steinberg, L., Fegley, S., & Dornbusch, S. M. (1993). Negative impact of part-time

work on adolescent adjustment: Evidence from a longitudinal study. *Developmental Psychology, 29,* 171–180.

Sternberg, R. J. (Ed.). (1999). *The nature of cognition.* Cambridge, MA: MIT Press.

Sternberg, R. J. (Ed.). (2000). *Handbook of intelligence.* New York: Cambridge University Press.

Sternberg, R. J. (Ed.). (2002). *Why smart people can be so stupid.* New Haven, CT: Yale University Press.

Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology, 49,* 479–502.

Sternberg, R. J., & Spear-Swerling, L. (Eds.). (1999). *Perspectives on learning disabilities* (pp. 3–21). New York: Westview/HarperCollins.

Sternberg, R. J., & Wagner, R. K. (Eds.). (1994). *Mind in context.* Cambridge, England: Cambridge University Press.

Stevenson, H., Stigler, J., Lee, S., Lucker, G., Kitamura, S., & Hsu, C. (1985). Cognitive performance and academic achievement of Japanese, Chinese, and American children. *Child Development, 56,* 718–734.

Stout, M. (2000). *The feel-good curriculum: The dumbing down of America's kids in the name of self-esteem.* Cambridge, MA: Perseus Publishing.

Strupp, H. H. (1989). Psychotherapy: Can the practitioner learn from the researcher? *American Psychologist, 44,* 717–724.

Strayer, D. L., & Johnston, W. A. (2001). Driven to distraction: Dual-task studies of simulated driving and and conversing on a cellular telephone. *Pschological Science, 12,* 462–466.

Sutherland, S. (1992). *Irrationality: The enemy within.* London: Constable.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47,* 143–148.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1,* 1–26.

Taleb, N. N. (2001). *Fooled by randomness.* New York: Texere.

Tassoni, C. J. (1996). Representativeness in the market for bets on National Football League games. *Journal of Behavioral Decision Making, 9,* 115–124.

Taube, K. T., & Linden, K. W. (1989). State mean SAT score as a function of participation rate and other educational and demographic variables. *Applied Measurement in Education, 2,* 143–159.

Teigen, K. H. (1986). Old truths or fresh insights? A study of students' evaluations of proverbs. *British Journal of Social Psychology, 25,* 43–49.

Thaler, R. H. (1992). *The winner's curse: Paradoxes and anomalies of economic life.* New York: Free Press.

Thomas, G., Alexander, K., & Eckland, B. (1979). Access to higher education: The importance of race, sex, social class, and academic credentials. *School Review, 87,* 133–156.

Thornton, E. (1986). *The Freudian fallacy.* London: Paladin Books.

Tooby, J. (2002, October 6). The greatest Englishman since Newton. *New York Times Book Review,* p. 12.

Turing, A. (1950). Computing machinery and intelligence. *Mind, 59,* 433–460.

Turner, C., Simons, L., Berkowitz, L., & Frodi, A. (1977). The stimulating and inhibiting effects of weapons on aggressive behavior. *Aggressive Behavior, 3,* 355–378.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Twachtman-Cullen, D. (1997). *A passion to believe.* Boulder, CO: Westview.

U.S. Congress House Select Committee on Aging. (1984, May 31). *Quackery: A $10 billion scandal.* Washington, DC: U.S. Government Printing Office.

University of California, Berkeley. (1991, January). The 18-year gap. *Berkeley Wellness Letter,* p. 2.

Updegrave, W. L. (1995, August). Why funds don't do better. *Money Magazine,* pp. 58–65.

Vallone, R., Griffin, D. W., Lin, S., & Ross, L. (1990). Overconfident prediction of future actions and outcomes by self and others. *Journal of Personality and Social Psychology, 58,* 582–592.

Vaughan, E. D. (1977). Misconceptions about psychology among introductory psychology students. *Teaching of Psychology, 4,* 138–141.

Violato, C. (1984). Effects of Canadianization of American-biased items on the WAIS

and WAIS-R information subtests. *Canadian Journal of Behavioral Science, 16,* 36–41.

Violato, C. (1986). Canadian versions of the information subtests of the Wechsler tests of intelligence. *Canadian Psychology, 27,* 69–74.

Wachs, T. D. (2000). *Necessary but not sufficient: The respective roles of single and multiple influences on individual development.* Washington, DC: American Psychological Association.

Wagenaar, W. A. (1988). *Paradoxes of gambling behavior.* Hove, England: Erlbaum.

Wagenaar, W. A., & Keren, G. (1986). The seat belt paradox: Effect of adopted roles on information seeking. *Organizational Behavior and Human Decision Processes, 38,* 1–6.

Wainer, H. (1989). Eelworms, bullet holes, and Geraldine Ferraro: Some problems with statistical adjustment and some solutions. *Journal of Educational Statistics, 14,* 121–140.

Wainer, H. (1993). Does spending money on education help? A reaction to the Heritage Foundation and the Wall Street Journal. *Educational Researcher, 22*(9), 22–24.

Wainer, H. (1999). The most dangerous profession: A note on nonsampling error. *Psychological Methods, 4,* 250–256.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes." *Psychological Bulletin, 122,* 203–215.

Watters, E., & Ofshe, R. (1999). *Therapy's delusions: The myth of the unconscious and the exploitation of the walking worried.* New York: Scribners.

Welch, D. A. (2002). *Decisions, decisions: The art of effective decision making.* Amherst, NY: Prometheus Books.

Wheeler, D. L., Jacobson, J. W., Paglieri, R. A., & Schwartz, A. A. (1993). An experimental assessment of facilitated communication. *Mental Retardation, 31,* 49–60.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: HarperCollins.

Winter, M. G., North, C., & Sugar, L. A. (2001). Psychologists' response to criticisms about research based on undergraduate participants: A developmental perspective. *Canadian Psychology, 42,* 216–225.

Wilson, E. O. (1998). *Consilience: The unity of knowledge.* New York: Knopf.

Wilson, R. A., & Keil, F. C. (Eds.). (1999). *The MIT Encyclopedia of the cognitive sciences.* Cambridge, MA: MIT Press.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117–142.

Wiseman, R., Beloff, J., & Morris, R. L. (1996). Testing the ESP claims of SORRAT. *Skeptical Inquirer, 20*(5), 45–61.

Wolfle, L. M. (1987). Enduring cognitive effects of public and private schools. *Educational Researcher, 16,* 5–11.

Wood, J. M., Nezworski, T., & Stejskal, W. J. (1996). The comprehensive system for the Rorschach: A critical examination. *Psychological Science, 7,* 3–10.

Wood, W., Wong, F. Y., & Chachere, J. G. (1991). Effects of media violence on viewers' aggression in unconstrained social interaction. *Psychological Bulletin, 109,* 371–383.

Woodcock, R. W. (1987). *Woodcock Reading Mastery Tests—Revised.* Circle Pines, MN: American Guidance Service.

Woolley, B. (2000). *The bride of science: Romance, reason, and Byron's daughter.* New York: McGraw-Hill.

Wright, L. (1988). The Type A behavior pattern and coronary artery disease. *American Psychologist, 43,* 2–14.

Wright, R. (1988). *Three scientists and their gods.* New York: Harper & Row.

Wright, R. (1999, December 13). The accidental creationist. *The New Yorker,* pp. 56–65.

Wrightsman, L. S. (1999). *Judicial decision making: Is psychology relevant?* New York: Kluwer Academic/Plenum.

Yates, J. F., Lee, J., & Bush, J. G. (1997). General knowledge overconfidence: Cross-national variations, response style, and "reality." *Organizational Behavior and Human Decision Processes, 70,* 87–94.

Yates, J. F., Lee, J., & Shinotsuka, H. (1996). Beliefs about overconfidence, including its cross-national variation. *Organizational Behavior and Human Decision Processes, 65,* 138–147.

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analysis comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin, 124,* 283–307.

Zechmeister, J. S., & Zechmeister, E. B. (2000). Introductory textbooks and psychology's core concepts. *Teaching of Psychology, 27,* 6–11.

Zill, N., & Winglee, M. (1990). *Who reads literature?* Cabin John, MD: Seven Locks Press.

Zweig, J. (1998, July). Here's how to use the news and tune out the noise. *Money Magazine,* pp. 63–64.

Zweig, J. (2001, May). Do you sabotage yourself? *Money Magazine,* pp. 75–78.

Zweig, J. (2002, October). Are you wired for wealth? *Money Magazine,* pp. 75–83.